Mat rials and Methods Involving Conditional Aggregation Domains

Background of the Invention

Various systems have been devised for ligand-dependent regulation of biological events including gene transcription, protein localization, induction of intracellular signaling, etc. These systems are based on a number of mechanisms, including ligand-induced allosteric changes in a fusion protein leading to transcription or repression of a target gene (systems based using tet, RU486, ecdysone, etc or analogs thereof) and ligand-induced crosslinking of fusion proteins leading to transcription or repression of a target gene, to induction of intracellular signaling, to protein localization and other biological actions (see e.g. WO 94/18317, 95/02684, 96/06097, 97/31899, 97/31898, 96/41865, 95/33052 and PCT/US98/17723 (dkt 363-C-PCT)). See also, Clackson, "Controlling mammalian gene expression with small molecules" Current Opinion in Chemical Biology, 1:210-218, 1997.

The subject invention provides a new system for regulating biological events.

Summary of the Invention

This invention takes a unique approach to regulation of biological events in cells, including transcription of a target gene, localization of a protein to a desired cellular compartment or site, etc. Compositions and methods of this invention are useful in biological research, heterologous gene expression and gene therapy applications.

Key features of the invention include conditional self-aggregating domains ("CADs"), fusion proteins containing them, ligands which bind to the CADs and permit the release or disaggregation of the fusion proteins, recombinant nucleic acids encoding such fusion proteins, compositions comprising one or more such recombinant nucleic acids (optionally together with one or more recombinant nucleic acids encoding accessory fusion proteins and/or a target gene construct), vectors containing such recombinant nucleic acids, cells transduced with these vectors and other material and important methods involving such. Key fusion proteins of the invention contain at least two mutually heterologous domains, one of which being a CAD. In some cases the fusion protein contains two or more CADs.

Proteins which contain one or more CADs aggregate with one another (i.e., self-aggregate) to form complexes containing two or more protein molecules. Following exposure to a ligand which binds to the CAD, the complexes disaggregate to liberate molecules of the CAD-containing fusion protein(s).

The practitioner may use one of the CADs disclosed in detail in this document, or may select one of his or her own choosing. We currently prefer CAds derived from immunophilin or cyclophilin domains, and in particular CADs derived from FKBPs such as human FKBP12. Particularly preferred CADs include domains containing peptide sequence derived from human FKBP12 in which one to

three amino acids have been replaced with independently selected different amino acids. Examples include FKBP12-derived domains in which either F36 or W59 is replaced with a different amino acid. Specific examples include F36M and W59V FKBP12 domains. Again, the practitioner may use any self-aggregating, ligand-dispersible domain in the design of fusion proteins for use in the practice of

5   this invention.

The portion of the fusion protein which is heterologous to the CAD may comprise any protein or protein domain of interest to the practitioner. For instance, the heterologous portion may comprise a DNA binding domain (such as a naturally occurring example like the GAL4 DNA binding domain, or a composite DNA binding domain such as a zinc finger composite or a ZFHD1 composite DNA binding

10   domain to recognize a DNA sequence which occurs naturally in the engineered cells or not), a transcription activation domain (such as a VP16 or p65 transcription activation domain), a transcription repression domain (such as a KRAB domain), a cellular localization domain (such as a membrane targeting domain, a membrane spanning domain or a myristoylation site; a nuclear target domain or a mitochondrial targeting domain) or a cellular signaling domain (such as a tyrosine kinase or other

15   signaling domain of a growth factor or cytokine receptor).

To illustrate the invention, in some cases a CAD-containing fusion protein contains a transcription regulating domain (e.g. a transcription activation or repression domain) and a DNA binding domain. Following exposure to the ligand for the CAD to liberate the fusion protein from its self-aggregated complex, the fusion protein is capable of activating or repressing the transcription of a

20   target gene construct containing a target gene in operative asssociation with a DNA sequence recognized by the DNA binding domain.

In other cases two CAD-containing fusion proteins are expressed in the cells. One of the fusion proteins contains a transcription regulating domain, the other a DNA binding domain which recognizes a DNA sequence operably linked to a target gene in the cells. In the absence of ligand, the

25   CAD-containing fusion proteins form a complex to activate or repress the transcription of the target gene, as the case may be. Following addition of the ligand for the CAD, the complex is dispelled and the transcriptional regulation is terminated.

In another example, two CAD-containing fusion proteins are expressed in the cells. One of the fusion proteins contains a transcription repression domain, the other a DNA binding domain which

30   recognizes a DNA sequence operably linked to a target gene in the cells. In the absence of ligand, the CAD-containing fusion proteins form a complex and repress the transcription of the target gene. Following addition of the ligand for the CAD, the complex is dispelled and the transcriptional regulation is terminated. The cells may further express a third fusion protein which comprises a transcription activation domain and at least one ligand binding domain. Addition of a ligand which is capable of

35   forming a complex containing the DNA binding domain-containing fusion protein and the third fusion protein (which contains the transcription activation domain) will activate the transcription of the target gene, formerly repressed. A special case is provided where the CAD is an FKBP-derived domain, the ligand binding domain of the third fusion protein is an FRB-derived domain, and the ligand for the CAD

(for example, a rapamycin analog or derivative) is also capable of forming a complex with the third fusion protein and the DNA binding domain-containing fusion protein. In that case, addition of ligand releases the transcription-repression fusion protein and replaces it with one or more transcription activating fusion proteins.

5   In another example, the CAD-containing fusion protein also contains a membrane-targeting domain and a cellular signaling domain (e.g. the cytoplasmic domain of a receptor for a growth factor or cytokine). These fusion proteins are designed to localize at the cell membrane, aggregate with one another and induce the cellular signal characteristic of the signaling domain. Addition of ligand dispells the complex and decreases or blocks continued signaling.

10   In another example, two CAD-containing fusion proteins are expressed in the cells. One of these fusion proteins contains one or more CADs and a membrane targeting domain. The other contains one or more CADs and a signaling domain whose signaling activity requires membrane anchoring. The first fusion protein is designed to recruit the second fusion protein to the cell membrane where it can signal. Addition of ligand dispells the complex and decreases or blocks

15   continued signaling.

Because the system is modular, the practitioner can readily adjust the design and configuration for a variety of applications.

One object of the invention is thus the fusion proteins described herein.

Another object of the invention is the recombinant nucleic acids encoding such fusion

20   proteins. Those recombinant nucleic acids may be operably linked to an expression control sequence permitting their expression in host cells into which they have been transduced.

Another object is a vector containing a recombinant nucleic acid of the invention, generally operably linked to an expression control sequence. Such vectors include "viral" vectors which contain part or all of a viral genome in addition to the recombinant nucleic acid encoding the fusion protein of

25   this invention. Viral vectors can be designed and used for the production of recombinant viruses harboring a recombinant nucleic acid of this invention. A wider variety of such viral systems are known in the art and may be adapted to the practice of this invention, including e.g. adenovirus, AAV, retrovirus, hybrid adeno-AAV, lentivirus and others.

Recombinant nucleic acids of this invention may be transduced into host cells by any available

30   means in order to render those cells capable of regulated production of a target protein. The cells are preferably eukaryotic cells, generally are animal cells, and in many embodiments are mammalian, whether human or non-human. The cells may be transduced in situ within their host organism, or they may be transduced while being maintained in vitro. The cells may be primary cells or may be from a cell line.

35   The invention thus provides methods for rendering a cell capable of ligand-dependent transcriptional regulation of a target gene which involves introducing into the cell a recombinant nucleic acid(s) of this invention to yield engineered cells which can express the encoded CAD-containing fusion protein(s) as described above. The recombinant nucleic acid(s) may be introduced

3

in viral or other form into cells maintained in vitro or into cells present within an organism. The resultant engineered cells and their progeny containing one or more of these recombinant nucleic acids may be used in a variety of important applications discussed elsewhere, including human gene therapy, analogous veterinary applications, the creation of cellular or animal models (including transgenic

5 applications), assay applications, and the production of a desired protein in vitro, e.g. for recovery and use. Such cells are useful, for example, in methods involving the addition of a ligand, preferably a cell permeant ligand, to the cells (or administration of the ligand to an organism containing the cells) to regulate (i.e., activate or repress) transcription of a target gene. Particularly important animal models include rodent (especially mouse and rat) and non-human primate models. In human gene therapy

10 applications, the cells will generally be human and the peptide sequence of each of the various domains present in the fusion proteins will preferably be, or be derived from, a peptide sequence of human origin, to the extent possible.

The invention also provides for nucleic acid compositions comprising two or more nucleic acids encoding fusion proteins, one or both of which contain CADs. One specialized application of

15 this invention involves the use of two recombinant nucleic acids, one encoding a first fusion protein comprising at least one CAD and a DNA binding domain, the other encoding a second fusion protein comprising at least one CAD and a transcription repression domain. A cell transduced with those recombinant nucleic acids which further contains a target gene operably linked to an expression control sequence to which the DNA binding domain binds, is useful for regulated expression of the

20 target gene. The two fusion proteins are designed to aggregate in the absence of ligand, and in so doing, recruit the repression domain to the expression control sequence of the target gene, thereby repressing its transcription. To regulatably express the target gene, one administers ligand to the cells or organism containing the cells.

Another object of the invention is thus a cell containing a first fusion protein comprising a CAD

25 and a DNA binding domain and a second fusion protein comprising a CAD and a transcription repression domain. The cell may further comprise a target gene operably linked to an expression control sequence to which the DNA binding domain binds.

Another object of the invention is an animal containing engineered cells as described herein.

30 **Brief Description of the Drawings**

Figure 1 illustrates six different formats for the use of CADs to control intracellular processes. Note that some F36M-FKBP moieties are shaded differently for clarity; this does not indicate any differences between the shaded and non-shaded molecules. (A) Use as a transcriptional "off-switch". (B) Use to

35 actively repress a gene by delivering repression domains to a promoter; then subsequent rreplacement with an activation domain to activate transcription. (C) Inactivation of a protein (here a transcription factor) by sequestering it through aggregation, until ligand is added. (D) A signaling "off-switch". Signaling is constitutively on until addition of ligand. (E) Conditional recruitment of a signaling

4

or other domain to the membrane. (F) Replacing the interaction domains of two proteins with CADs to allow their interaction to be disrupted by addition of CAD ligand.

Figure 2: Ligand-reversible self-association of F36M fusion proteins was identified using a mammalian two-hybrid assay.

Fig. 2A: Wild-type (W), F36V-FKBP (V) or F36M-FKBP (M) were fused in one or three copies to the composite DNA binding domain ZFHD1 and separately to the activation domain of human p65 (amino acids 361-550), and co-transfected into mammalian cells for interaction analysis, as described (Pollock and Rivera, Meth. Enz. 306:263-281, 1999). Only F36M fusion proteins interacted in the absence of dimerizer drug (AP1510). Z-p65 indicates a covalent fusion of ZFHD1 and p65.

Fig. 2B: Transcription resulting from F36M-F36M interactions is inhibited to baseline levels by monomeric FKBP ligands such as FK506, AP219988 and 22542, but not the negative control compound cyclosporin A (CsA) that has no affinity for FKBPs.

Figure 3 shows dose-dependent target gene transcription following addition of ligand (FK506) for the F36M FKBP CAD. ZFHD-p65 is a positive control which is not affected by the addition of ligand.

Figure 4 shows dose-dependent inhibition of target gene transcription following addition of ligand (FK506) for the W59V FKBP CAD.

Figure 5 illustrates a construct for use in screening to identify novel CADs.

Detailed Description

Definitions:

For convenience, the intended meaning of certain terms and phrases used herein are provided below.

"Capable of selectively hybridizing" means that two DNA molecules are susceptible to detectable hybridization with one another, despite the presence of other DNA molecules, under hybridization conditions which can be chosen or readily determined empirically by the practitioner of ordinary skill in this art. Such treatments include conditions of high stringency such as washing extensively with buffers containing 0.2 to 6 x SSC, and/or containing 0.1% to 1% SDS, at temperatures ranging from room temperature to 65-75°C. See for example F.M. Ausubel et al., Eds, Short Protocols in Molecular Biology, Units 6.3 and 6.4 (John Wiley and Sons, New York, 3d Edition, 1995).

"Cells", "host cells" or **"recombinant host cells"** refer not only to the particular cells under discussion, but also to their progeny. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be

identical to the parent cell, but are still included within the scope of the term as used herein.

"**Cell line**" refers to a population of cells capable of continuous or prolonged growth and division in vitro. Often, cell lines are clonal populations derived from a single progenitor cell. It is further known in the art that spontaneous or induced changes can occur in karyotype during storage or

5   transfer of such clonal populations. Therefore, cells derived from a given cell line may not be precisely identical to the ancestral cells or cultures, and the cell line referred to includes such variants.

"**Composite**", "**fusion**", and "**recombinant**" denote a material such as a nucleic acid, nucleic acid sequence or polypeptide which contains at least two constituent portions which are mutually heterologous in the sense that they are not otherwise found directly (covalently) linked in

10   nature, i.e., are not found in the same continuous polypeptide or gene in nature, at least not in the same order or orientation or with the same spacing present in the composite, fusion or recombinant product. Typically, such materials contain components derived from at least two different proteins or genes or from at least two non-adjacent portions of the same protein or gene. In general, "composite" refers to portions of different proteins or nucleic acids which are joined together to form a single

15   functional unit, while "fusion" generally refers to two or more functional units which are linked together. "Recombinant" is generally used in the context of nucleic acids or nucleic acid sequences.

A "**coding sequence**" or a sequence which "encodes" a particular polypeptide or RNA, is a nucleic acid sequence which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide in vitro or in vivo when placed under the control of an appropriate expression

20   control sequence. The boundaries of the coding sequence are generally determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from procaryotic or eukaryotic mRNA, genomic DNA sequences from procaryotic or eukaryotic DNA, and synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the coding sequence.

25   A "**construct**", e.g., a "nucleic acid construct" or "DNA construct", refers to a nucleic acid or nucleic acid sequence.

"**Derived from**" denotes a peptide or nucleotide sequence selected from within a given sequence. A peptide or nucleotide sequence derived from a named sequence may further contain a small number of modifications relative to the parent sequence, in most cases representing deletion,

30   replacement or insertion of less than about 15%, preferably less than about 10%, and in many cases less than about 5%, of amino acid residues or bases present in the parent sequence. In the case of DNAs, one DNA molecule is also considered to be derived from another if the two are capable of selectively hybridizing to one another. Polypeptides or polypeptide sequences are also considered to be derived from a reference polypeptide or polypeptide sequence if any DNAs encoding the two

35   polypeptides or sequences are capable of selectively hybridizing to one another. Typically, a derived peptide sequence will differ from a parent sequence by the replacement of up to 5 amino acids, in many cases up to 3 amino acids, and very often by 0 or 1 amino acids. A derived nucleic acid sequence will differ from a parent sequence by the replacement of up to 15 bases, in many cases up

to 9 bases, and very often by 0 - 3 bases. In some cases the amino acid(s) or base(s) is/are added or deleted rather than replaced.

"D main" refers to a portion of a protein or polypeptide. In the art, the term "domain" may refer to a portion of a protein having a discrete secondary structure. However, as will be apparent from the context used herein, the term "domain" as used in this document does not necessarily connote a given secondary structure. Rather, a peptide sequence is referred to herein as a "domain" simply to denote a polypeptide sequence from a defined source, or having or conferring an intended or observed activity. Domains can be derived from naturally occurring proteins or may comprise non-naturally-occurring sequence.

"Expression control element", or simply "control element", refers to DNA sequences, such as initiation signals, enhancers, promoters and silencers, which induce or control transcription of DNA sequences with which they are operably linked. Control elements of a gene may be located in introns, exons, coding regions, and 3' flanking sequences. Some control elements are "tissue specific", i.e., affect expression of the selected DNA sequence preferentially in specific cells (e.g., cells of a specific tissue), while others are active in many or most cell types. Gene expression occurs preferentially in a specific cell if expression in this cell type is observably higher than expression in other cell types. Control elements include so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well. Furthermore, a control element can act constitutively or inducibly. An inducible promoter, for example, is demonstrably more active in response to a stimulus than in the absence of that stimulus. A stimulus can comprise a hormone, cytokine, heavy metal, phorbol ester, cyclic AMP (cAMP), retinoic acid or derivative thereof, etc. A nucleotide sequence containing one or more expression control elements may be referred to as an "expression control sequence".

"Gene" refers to a nucleic acid molecule or sequence comprising an open reading frame and including at least one exon and (optionally) one or more intron sequences.

"Genetically engineered cells" denotes cells which have been modified by the introduction of recombinant or heterologous nucleic acids (e.g. one or more DNA constructs or their RNA counterparts) and further includes the progeny of such cells which retain part or all of such genetic modification.

"Heterologous", as it relates to nucleic acid or peptide sequences, denotes sequences that are not normally joined together, and/or are not normally associated with a particular cell. Thus, a "heterologous" region of a nucleic acid construct is a segment of nucleic acid within or attached to another nucleic acid molecule that is not found in association with the other molecule in nature. For example, a heterologous region of a construct could include a coding sequence flanked by sequences not found in association with the coding sequence in nature. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., synthetic sequences having codons different from the native gene). Similarly, in the case of a cell transduced with a nucleic acid construct which is not normally present in the cell, the cell and the

construct would be considered mutually heterologous for purposes of this invention.

"Interact" refers to directly or indirectly detectable interactions between molecules, such as can be detected using, for example, a yeast two hybrid assay or by immunoprecipitation. The term "interact" encompasses "binding" interactions between molecules. Interactions may be, for example,
5   protein-protein, protein-nucleic acid, protein-small molecule or small molecule-nucleic acid in nature.

"Nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include derivatives, variants and analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

10   A "polylinker", also sometimes referred to as a "multiple cloning site" is a region within a vector which contains multiple sites for restriction enzyme cleavage, thus rendering the vector suitable for cloning of exogenous genes.

"Protein", "polypeptide" and "peptide" are used interchangeably.

A "recombinant virus" is a virus particle in which the packaged nucleic acid contains a
15   heterologous portion.

The "secretory machinery" of the cell refers to the cellular compartments to which secreted and membrane proteins are targeted and processed. These compartments include the endoplasmic reticulum (ER) and the cis, medial and trans Golgi. In this document, the term ER is often used generically to mean "secretory compartment."

20   A "target gene" is a nucleic acid of interest, the transcription of which is modulated according to the methods of the invention. The target gene can encode, for instance, a protein, an antisense RNA or a ribozyme.

A "therapeutic protein" is a protein of interest, the production of which is modulated according to the methods of the invention. The therapeutic protein can be, for example, a hormone, an
25   endorphin, etc.

"Transfection" means the introduction of a naked nucleic acid molecule into a recipient cell. "Infection" refers to the process wherein a nucleic acid is introduced into a cell by a virus containing that nucleic acid. A "productive infection" refers to the process wherein a virus enters the cell, is replicated, and is then released from the cell (sometimes referred to as a "lytic" infection).
30   "Transduction" encompasses the introduction of nucleic acid into cells by any means.

"Transgene" refers to a nucleic acid sequence which has been introduced into a cell. Daughter cells deriving from a cell in which a transgene has been introduced are also said to contain the transgene (unless it has been deleted). The polypeptide or RNA encoded by a transgene may be partly or entirely heterologous, i.e., foreign, with respect to the animal or cell into which it is introduced.
35   Alternatively, the transgene can be homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene). A transgene can also be present in an episome. A

8

transgene can include one or more expression control elements and any other nucleic acid, (e.g. intron), that may be necessary or desirable for optimal expression of a selected coding sequence.

The term "v ctor" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is an episome, i.e., a nucleic acid capable of extra-
5   chromosomal replication. Often vectors are used which are capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of an included gene operatively linked to an expression control sequence can be referred to as "expression vectors". Expression vectors are typically in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the
10   present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of vectors which serve equivalent functions and which are or become known in the art. Viral vectors are nucleic acid molecules containing viral sequences which can be packaged into viral particles.

15   **CADs**

Fusion proteins containing one or more "conditional aggregation domains" (CADs) form aggregates with one another which are dispersed in the presence of ligand. Fusion proteins containing CADs are retained in cellular compartments, e.g. the cytoplasm or the nucleus. Such fusion proteins can also have nuclear localization sequences, which target the aggregates to the
20   nucleus.

In a preferred embodiment, the CAD is derived from human FKBP12. In particular, the FKBP mutant F36M functions as a conditional aggregation domain when fused to a heterologous target sequence in eukaryotic, e.g. mammalian, cells. In the absence of ligand, fusion proteins containing FKBP F36M self-aggregate and accumulate in complexes. Upon addition of ligand, the fusion protein
25   disaggregates. Another FKBP mutant which functions as a CAD is FKBP W59V (see example 10).

Ligands for CADs:

A wide variety of ligands, including both naturally occurring and synthetic substances, can be used in this invention to effect disaggregation of the fusion protein molecules. Criteria for selecting a
30   ligand are: (A) physiologic acceptability of the ligand (i.e., the ligand lacks undue toxicity towards the cell or animal for which it is to be used), (B) reasonable therapeutic dosage range, (C) suitability for oral administration (i.e., suitable stability in the gastrointestinal system and absorption into the vascular system), for applications in whole animals, including gene therapy applications, (D) ability to cross cellular and other membranes, as necessary, and (E) reasonable binding affinity for the CAD (for the
35   desired application). Preferably the compound is relatively physiologically inert, but for its affinity for the CAD. The less the ligand binds to native proteins or other materials within the cells to be targeted, the better the response will normally be. Preferably the ligand will be other than a peptide or nucleic acid, and will preferably have a molecular weight of less than about 5000 Daltons, more preferably less

than about 1200 Daltons.

In various embodiments where a ligand binding domain for a candidate ligand is endogenous to the cells to be engineered, it is often desirable to alter the peptide sequence of the ligand binding domain and to use a ligand which discriminates between the endogenous and engineered ligand

5   binding domains. Such a ligand should bind preferentially to the engineered ligand binding domain relative to a naturally occurring peptide sequence, e.g., from which the modified domain was derived. This approach can avoid untoward intrinsic activities of the ligand. Significant guidance and illustrative examples toward that end are provided in the various references cited herein.

Substantial structural modification of a ligand for a ligand binding domain is permitted, so long

10   as the modified compound still functions as a ligand for the ligand binding domain of interest, i.e., so long as the compound possesses sufficient binding affinity and specificity to function as disclosed herein. Some of the compounds will be macrocyclics, e.g. macrolides, although linear and branched compounds may be preferred in specific embodiments. Suitable binding affinities will be reflected in Kd values well below 10^{-4}, preferably below 10^{-6}, more preferably below about 10^{-7}, although binding

15   affinities below 10^{-9} or 10^{-10} are possible, and in some cases will be most desirable.

Illustrative examples of ligand binding domain/ligand pairs include retinol binding protein or variants thereof and retinol or derivatives thereof; cyclophilin or variants thereof and cyclosporin or analogs thereof; FKBP or variants thereof and FK506, FK520, rapamycin, analogs thereof or synthetic FKBP ligands. In the case of a ligand binding domain comprising or derived from an

20   immunophilin or cyclophilin, the complex of the ligand with the ligand binding domain will desirably not bind specifically to calcineurin or FRAP. A wide variety of FK506 derivatives and synthetic FKBP ligands are known which do not have observable immunosuppressive activity. Likewise, a variety of rapamycin analogs are known which bind to FKBP but are not immunosuppressive. See e.g. WO 98/02441for non-immunosuppressive rapalogs. Those and other ligands can be used as well,

25   depending on the choice of CAD.   Numerous assays are known in the art for identifying ligands which bind to CADs that are identified through screening, as described below.

Ligand binding domain/ligand pairs are illustrated by FKBP domains, e.g. F36M FKBP, and FKBP ligands. In general, it is preferred that the ligand bind preferentially to a mutated (i.e., having a peptide sequence not naturally occurring in the cells to be engineered) FKBP relative to wild-type

30   FKBP. Ligands for FKBP proteins, including F36M FKBP, can comprise or be derived from a naturally occurring FKBP ligand such as rapamycin, FK506 or FK520, or a synthetic FKBP ligand, e.g. as disclosed in PCT/US95/10559; Holt, et al., J. Amer. Chem. Soc.,1993, 115, 9925-9938; Holt, et al., Biomed. Chem. Lett., 1993, 4, 315-320; Luengo, et al., Biomed. Chem. Lett., 1993, 4, 321-324; Yamashita, et al., Biomed. Chem. Lett., 1993, 4, 325-328; PCT/US94/01617; PCT/US94/08008.

35   See also EP 0 455 427 A1; EP 0 465 426 A1; US 5,023,26; WO 92/00278; WO 94/18317; WO 97/31898; WO 96/41865; and Van Duyne et al (1991) Science 252, 839.

Illustrative types of ligands for FKBP-derived ligand binding domains include the following Genus I:
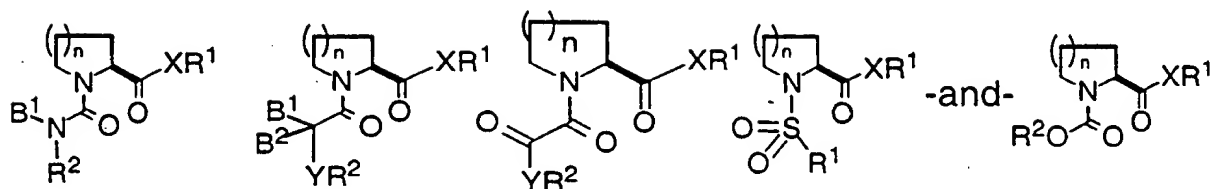
where

n = 1 or 2;

X = O, S, NH or CH$_2$;

B^1 and B^2 are independently H or aliphatic, heteroaliphatic, aryl or heteroaryl as those terms are defined below, usually containing one to about 12 carbon atoms (not counting carbon atoms of optional substituents);

Y = O, S, NH, -NH(C=O)-, -NH(C=O)-O-, -NH(SO$_2$)- or NR^3, or represents a direct, i.e. covalent, bond from R^2 to carbon 9;

R^1, R^2, and R^3 are aliphatic, heteroaliphatic, aryl or heteroaryl, usually containing one to about 36 carbon atoms (not counting carbon atoms of optional substituents);

two or more of B^1, B^2 and R^2 may be covalently linked to form a C3-C7 cyclic or heterocyclic moiety; and,

The term "aliphatic" as used herein includes both saturated and unsaturated straight chain, branched, cyclic, or polycyclic aliphatic hydrocarbons, which are optionally substituted with one or more substituents.

The term "substituents" includes aliphatic, aryl, heteroaryl and heterocyclic moietites, which may themselves be substituted, as well as functional groups such as R^8, -OR^8, -SR^8, -CN, -CHO, =O, -COOH, -COR^8, OS(O)$_2$ R^8, -SO$_2$-NHR^8, -NHSO$_2$ R^8, sulfate, sulfonate, (or ester, carbamate, urea, oxime or carbonate thereof), -NH$_2$ (or substituted amine, amide, urea, carbamate or guanidino derivative therof), halo, trihaloalkyl, -SO$_2$-CF$_3$, and -OSO$_2$F, where R^8 may be H, aliphatic, aryl, heteroaryl or heteroaliphatic. Aliphatic, heteraliphatic, aryl and heterocyclic substituents may themselves be substituted or unsubstituted (e.g. mono-, di- and tri-alkoxyphenyl; methylenedioxyphenyl or ethylenedioxyphenyl; halophenyl; or -phenyl-C(Me)$_2$-CH$_2$-O-CO-[C3-C6] alkyl or alkylamino). Additional examples of substituents are illustrated by the specific embodiments shown in the Examples which follow. (Unless otherwise specified, the alkyl, other aliphatic, alkoxy and acyl groups preferably contain 1-8, and in many cases 1-6, contiguous aliphatic carbon atoms).

The term "aliphatic" is thus intended to include alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, and cycloalkynyl moieties.

As used herein, the term "alkyl" includes both straight and branched alkyl groups. An analogous convention applies to other generic terms such as "alkenyl", "alkynyl" and the like. Furthermore, as used herein, the language "alkyl", "alkenyl", "alkynyl" and the like encompasses both

substituted and unsubstituted groups.

The term "alkyl" refers to groups usually having one to eight, preferably one to six carbon atoms. For example, "alkyl" may refer to methyl, ethyl, n-propyl, isopropyl, butyl, isobutyl, sec-butyl, tert-butyl, pentyl, isopentyl tert-pentyl, hexyl, isohexyl, and the like. Suitable substituted alkyls include, but are not limited to, fluoromethyl, difluoromethyl, trifluoromethyl, 2-fluoroethyl, 3-fluoropropyl, hydroxymethyl, 2-hydroxyethyl, 3-hydroxypropyl, and the like.

The term "alkenyl" refers to groups usually having two to eight, preferably two to six carbon atoms. For example, "alkenyl" may refer to prop-2-enyl, but-2-enyl, but-3-enyl, 2-methylprop-2-enyl, hex-2-enyl, hex-5-enyl, 2,3-dimethylbut-2-enyl, and the like. The language "alkynyl," which also refers to groups having two to eight, preferably two to six carbons, includes, but is not limited to, prop-2-ynyl, but-2-ynyl, but-3-ynyl, pent-2-ynyl, 3-methylpent-4-ynyl, hex-2-ynyl, hex-5-ynyl, and the like.

The term "cycloalkyl" as used herein refers to groups having three to seven, preferably three to six carbon atoms. Suitable cycloalkyls include, but are not limited to cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl and the like.

The term "heteroaliphatic" as used herein refers to aliphatic moieties which contain one or more oxygen, sulfur, or nitrogen atoms, e.g., in place of carbon atoms.

The term "heterocycle" as used herein refers to cyclic aliphatic groups having one or more heteroatoms, and preferably three to seven ring atoms total, includes, but is not limited to oxetane, tetrahydrofuranyl, tetrahydropyranyl, aziridine, azetidine, pyrrolidine, piperidine, morpholine, piperazine and the like.

The terms "aryl" and "heteroaryl" as used herein refer to stable mono- or polycyclic, heterocyclic, polycyclic, and polyheterocyclic unsaturated moieties having 3 - 14 carbon atom which may be substituted or unsubstituted. Non-limiting examples of useful aryl ring groups include phenyl, halophenyl, alkoxyphenyl, dialkoxyphenyl, trialkoxyphenyl, alkylenedioxyphenyl, naphthyl, phenanthryl, anthryl, phenanthro and the like. Examples of typical heteroaryl rings include 5-membered monocyclic ring groups such as thienyl, pyrrolyl, imidazolyl, pyrazolyl, furyl, isothiazolyl, furazanyl, isoxazolyl, thiazolyl and the like; 6-membered monocyclic groups such as pyridyl, pyrazinyl, pyrimidinyl, pyridazinyl, triazinyl and the like; and polycyclic heterocyclic ring groups such as benzo[b]thienyl, naphtho[2,3-b]thienyl, thianthrenyl, isobenzofuranyl, chromenyl, xanthenyl, phenoxathienyl, indolizinyl, isoindolyl, indolyl, indazolyl, purinyl, isoquinolyl, quinolyl, phthalazinyl, naphthyridinyl, quinoxalinyl, quinazolinyl, benzothiazole, benzimidazole, tetrahydroquinoline cinnolinyl, pteridinyl, carbazolyl, beta-carbolinyl, phenanthridinyl, acridinyl, perimidinyl, phenanthrolinyl, phenazinyl, isothiazolyl, phenothiazinyl, phenoxazinyl, and the like(see e.g. Katritzky, Handbook of Heterocyclic Chemistry). The aryl or heteroaryl moieties may be substituted with one to five members selected from the group consisting of hydroxy, C1-C8 alkoxy, C1-C8 branched or straight-chain alkyl, acyloxy, carbamoyl, amino, N-acylamino, nitro, halo, trihalomethyl, cyano, and carboxyl.

A "halo" substituent according to the present invention may be a fluoro, chloro, bromo or iodo

substituent.

As discussed above, R^1 may be aliphatic, heteroaliphatic, aryl or heteroaryl and usually comprises one to about 36 carbon atoms, exclusive of optional substituents.

In certain embodiments, R^1 is optionally be joined, i.e., covalently linked, to R^2, B^1 or B^2,

5   forming a macrocyclic structure.

In certain embodiments $-XR^1$ is a moiety of the formula

$$R^4-\underset{\underset{\underset{X}{|}}{R^5}}{}\overset{R^6-R^7}{}$$
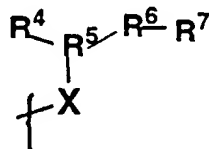
10   where R^4 is a H, aliphatic, heteroaliphatic, aryl or heteroaryl. The aliphatic moieties may be branched, unbranched, cyclic, saturated or unsaturated, substituted or unsubstituted and include, e.g, methyl, ethyl, isopropyl, t-butyl, cyclopentyl, cyclohexyl, etc. Heteroaliphatic moieties may be branched, unbranched or cyclic and include heterocycles such as morpholino, pyrrolidinyl, etc. Illustrative ortho-, meta- or para-, substitutents for a phenyl group at this position include one or more of the following:

15   halo, e.g. chloro or flouro; hydroxyl, amino, $-SO_2NH_2$, $-SO_2NH(aliphatic)$, $-SO_2N(aliphatic)_2$, $-O$-

aliphatic-COOH, $-O$-aliphatic-NH_2 (which may contain one or two N-aliphatic or N-acyl substituents),

C1-C6 alkyl, acyl, acyloxy, C1-C6 alkoxy, e.g. methoxy, ethoxy, methylenedioxy, ethylenedioxy, etc. Heteroaryl groups are as discussed previously, including indolyl, pyridyl, pyrrolyl, etc. Particular R^4 moieties include the following:

20

$z = 1-6$

$Q = -NH_2$, -NHalkyl, -Ndialkyl, -COOH, or -OH
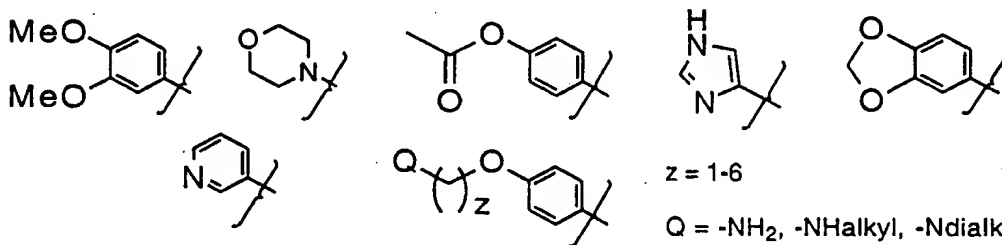
R^5 is a branched, unbranched or cyclic aliphatic moiety of 1 to 8 carbon atoms, which may be optionally substituted, including for example, -CH-, -CHCH2-, $-CH_2CH$-, $-CHCH_2CH_2$-, $-CH_2CHCH_2$-,-

25    $CH(CH_3)-CH_2-CH$, $-CH(CH_2CH_3)-CH_2-CH$, $-CH_2CH_2CH$-, $-C(CH_3)CH_2$-, and the like;

R^6 is an aliphatic, heteroaliphatic, heterocylic, aryl or heteroaryl moiety, which may be substituted or unsubstituted. Typical substituents for R^6 include branched, unbranched or cyclic, C1-C8, aliphatic or heteroaliphatic groups, including unsaturated groups such as substitute or unsubstituted alkenes, heterocycles, phenyl, etc.

30   R^7 is H or a substituent such as, in certain embodiments, $-(CH_2)_z-CH=CH_2$, $-(CH_2)_z-COOH$, -

$(CH_2)_z$-CHO, -$(CH_2)_z$-OH, -$(CH_2)_z$-NH$_2$, -$(CH_2)_z$-NH-alkyl, -$(CH_2)_z$-SH, or an amino group which may be substituted or unsubstituted (preferably a tertiary amine), etc. In embodiments where R^6 is aryl, R^7 may be present in the o, m, or p position. z is an integer from 0 through 4.

As discussed above, B^1, B^2 and R^2 may be aliphatic, heteroaliphatic, aryl or heteroaryl. Typical groups include a branched, unbranched or cyclic, saturated or unsaturated, aliphatic moiety, preferably of 1 to about 12 carbon atoms (including for example methyl, ethyl, n-propyl, isopropyl, cyclopropyl, -CH$_2$-cyclopropyl, allyl, n-butyl, sec-butyl, isobutyl, tert-butyl, cyclobutyl, -CH$_2$-cyclobutyl, n-pentyl, sec-pentyl, isopentyl, tert-pentyl, cyclopentyl, -CH$_2$-cyclopentyl, n-hexyl, sec-hexyl, cyclohexyl, -CH$_2$-cyclohexyl and the like), which aliphatic moiety may optionally be substituted with an-OH, -C=O, -COOH, CHO, allyl, NH$_2$ (or substituted amine, amide, urea or carbamate), ether (or thio-ether, in either case, aliphatic or aromatic), aryl, or heteroaryl moiety, and may optionally contain a heteroatom in place of one or more CH$_2$ or CH units; or a substituted or unsubstituted aryl (e.g. mono-, di- and tri-alkoxyphenyl; methylenedioxyphenyl or ethylenedioxyphenyl; halophenyl; or -phenyl-C(Me)$_2$-CH$_2$-O-CO-[C3-C6] alkyl or alkylamino) or heteroaromatic moiety. In such embodiments, where YR^2 is -OPhenyl and B^1 is H, B^2 is preferably not cyclopentyl. In other embodiments, Y is NH and the moiety -(C=O)-CH(B^1)NHR^2 comprises among other groups, D- or L-forms of naturally occurring or synthetic alpha amino acids as well as N-alkyl, N-acyl, N-aryl and N-aroyl derivatives thereof. Particular XR^1, G, B^1, B^2 and YR^2 groups for the various foregoing structures further include those illustrated in compounds described in the examples, tables of monomers and dimers and other disclosure in WO 96/06097, WO 97/31899 and WO 97/31898.

One preferred class of compounds are those compounds of Genus I in which n is 2.

Another preferred class of compounds are those compounds of Genus I in which B^1 is H; B^2 is branched, unbranched or cyclic, saturated or unsaturated, aliphatic moiety, preferably of 1 to 8, more preferably 1 to 6, carbon atoms (including for example methyl, ethyl, n-propyl, isopropyl, cyclopropyl,-CH$_2$-cyclopropyl, allyl, n-butyl, sec-butyl, isobutyl, tert-butyl, cyclobutyl, -CH$_2$-cyclobutyl, n-pentyl, sec-pentyl, isopentyl, tert-pentyl, cyclopentyl, -CH$_2$-cyclopentyl, n-hexyl, sec-hexyl, cyclohexyl, -CH$_2$-cyclohexyl and the like), which aliphatic moiety may optionally be substituted, e.g. with an -OH, -C=O,-COOH, CHO, allyl, NH$_2$ (or substituted amine, amide, urea or carbamate), or ether (or thio-ether, in either case, aliphatic or aromatic), and may optionally contain a heteroatom in place of one or more CH$_2$ or CH units; and YR^2 is aryl, heteroaryl and may be optionally substituted (YR^2, for instance, includes moieties such as o-, m-, or p-alkoxyphenyl; 3,5-, 2,3-, 2,4-, 2,5-, 3,4- or 3,5-dialkoxyphenyl, or 3,4,5-trialkoxyphenyl, e.g. where the alkoxy groups are independently selected from methoxy and ethoxy (one or more of which may bear a hydroxy or amino moiety).

Another preferred class of compounds are those compounds of Genus I in which B^1, B^2 and YR^2 are the same or different lower aliphatic moieties.

14

Another preferred class of compounds are those compounds of Genus I which contain a moiety $-NB^1R^2$ in which B^1 is H and R^2 is lower aliphatic.

Another preferred class of compound are those compounds of Genus I in which G is an alicyclic or heterocyclic group bearing optional substituents.

5   Another preferred class of compounds are those compounds of Genus I in which X is oxygen and R^1 comprises $R^4R^5R^6R^7$ where R^4 is aliphatic, alicyclic, aryl, heteroaryl, or heterocyclic, optionally substituted; R^5 is a branched or unbranched lower aliphatic group; R^6 is aliphatic, alicyclic, heteroaliphatic, heterocyclic, aryl or heteroaryl, optionally substituted.

Another preferred class of compounds are those compounds of Genus I in which R1

10   comprises $R^4R^5R^6R^7$ as described in the immediately preceding paragraph and YR^2comprises a substituted or unsubstituted aryl or heteroaryl, including phenyl; o-, m- or p- substituted phenyl where the substituent is halo such as chloro, lower alkyl, or alkoxy, such as methoxy or ethoxy; disubstituted phenyl, e.g. dialkoxyphenyl such as 2,4-, 3,4- or 3.5-dimethoxy or diethoxy phenyl or such as methylenedioxyphenyl, or 3-methoxy-5-ethoxyphenyl; or trisubstituted phenyl, such as trialkoxy

15   (e.g., 3,4,5-trimethoxy or ethoxyphenyl), 3,5-dimethoxy-4-chloro-phenyl, etc.).

In addition, such compounds may comprise a substituted proline and pipecolic acid derivative, numerous examples of which have been described in the literature. Using synthetic procedures similar to those described in the patent documents and scientitific literature cited herein, substituted prolines and pipecolates can be utilized to prepare ligands with substituents at positions C-2 to C-6

20   (with reference to the FK506 numbering of most of the references cited below), as exemplified in the patent applications cited herein.

For representative examples of substituted prolines and pipecolic acids see: Chung, et al., J. Org. Chem., 1990, 55, 270; Shuman, et al., J. Org. Chem., 1990, 55, 738; Hanson, et al., Tetrahedron Lett., 1989, 30, 5751; Bailey, et al., Tetrahedron Lett., 1989, 30, 6781.

25   For a variety of guidance on chemical transformations, synthesis, formulation and delivery of a variety of compounds, including additional information relating to FKBP ligands and/or to ligands for other ligand binding domains, see e.g., WO 94/18317 and Belshaw et al, 1996, PNAS 93:4604-4607) (for methods and materials based on ligands for an immunophilin such as FKBP, a cyclophilin, and/or FRB domain); WO 96/06097 and WO 97/31898 (more ligands for FKBP and variants thereof);

30   WO 93/33052, WO 96/41865 and Rivera et al, "A humanized system for pharmacologic control of gene expression", Nature Medicine 2(9):1028-1032 (1997)) (rapamycin analogs); WO 94/18317 (cyclophilin/cyclosporin); Licitra et al, 1996, Proc. Natl. Acad. Sci. USA 93:12817-12821 (DHFR/methotrexate); and Farrar et al, 1996, Nature 383:178-181 (DNA gyrase/coumermycin). Numerous variations and modifications to ligands and ligand binding domains, as well as

35   methodologies for designing, selecting and/or characterizing them, which may be adapted to the present invention are disclosed in the cited references.

Target genes:

Target genes whose transcription is regulated in various embodiments of this invention may endogenous or heterologous to the engineered cells. The target gene can be a membrane-bound or membrane-spanning protein, a secreted protein, or a cytoplasmic protein. The proteins which are expressed, singly or in combination, can involve homing, cytotoxicity, proliferation, differentiation,

5   immune response, inflammatory response, clotting, thrombolysis, hormonal regulation, angiogenesis, etc. The polypeptide may be of naturally occurring or non-naturally occurring peptide sequence.

Various secreted products include hormones, such as insulin, human growth hormone, glucagon, pituitary releasing factor, ACTH, melanotropin, relaxin, leptin, etc.; growth factors, such as EGF, IGF-1, TGF-alpha, -beta, PDGF, G-CSF, M-CSF, GM-CSF, members of the FGF family,

10   erythropoietin, thrombopoietin, megakaryocytic growth factors, nerve growth factors, etc.; proteins which stimulate or inhibit angiogenesis such as angiostatin, endostatin and VEGF and variants thereof; interleukins, such as IL-1 to -15; TNF-alpha and -beta; interferons -alpha, -beta and -gamma; and enzymes and other factors, such as tissue plasminogen activator, members of the complement cascade, perforins, superoxide dismutase; coagulation-related factors such as antithrombin-III, Factor

15   V, Factor VII, Factor VIIIc, vWF, Factor IX, alpha-anti-trypsin, protein C, and protein S; endorphins, dynorphin, bone morphogenetic protein, CFTR, etc.

The protein may be a naturally-occurring surface membrane protein or a protein made so by introduction of an appropriate signal peptide and transmembrane sequence. Various such proteins include homing receptors, e.g. L-selectin (Mel-14), hematopoietic cell markers, e.g. CD3, CD4, CD8,

20   B cell receptor, TCR subunits alpha, beta, gamma or delta, CD10, CD19, CD28, CD33, CD38, CD41, etc., receptors, such as the interleukin receptors IL-2R, IL-4R, etc.; receptors for other ligands including the various hormones, growth factors, etc.; receptor antagonists for such receptors and soluble forms of such receptors; channel proteins, for influx or efflux of ions, e.g. H+, Ca+2, K+, Na+, Cl-, etc., and the like; CFTR, tyrosine activation motif, zap-70, etc.

25   The target protein can be an intracellular protein such as a protein involved in a metabolic pathway, or a regulatory protein, steroid receptor, transcription factor, etc.,

By way of further illustration, in T-cells, one may wish to introduce genes encoding one or both chains of a T-cell receptor. For B-cells, one could provide the heavy and light chains for an immunoglobulin for secretion. For cutaneous cells, e.g. keratinocytes, particularly keratinocyte stem

30   cells , one could provide for protection against infection, by secreting alpha, beta or gamma interferon, antichemotactic factors, proteases specific for bacterial cell wall proteins, various anti-viral proteins, etc.

In various situations, one may wish to direct a cell to a particular site. The site can include anatomical sites, such as lymph nodes, mucosal tissue, skin, synovium, lung or other internal organs or functional sites, such as clots, injured sites, sites of surgical manipulation, inflammation, infection,

35   etc. Regulated expression of a membrane protein which recognizes or binds to the particular site of interest, for example, provides a method for directing the engineered cells to that site. Thus one can achieve a localized concentration of a secreted product or effect cell-based healing, scavenging, protection from infection, anti-tumor activity, etc. Proteins of interest include homing receptors, e.g. L-

selectin, GMP140, CLAM-1, *etc.*, or addr ssins, *e.g.* ELAM-1, PNAd, LNAd, *etc.*, clot binding proteins, or cell surface proteins that respond to localized gradients of chemotactic factors.

In one embodiment of this invention, binding of a ligand to a CAD regulates transcription of a target gene. In this embodiment, the target gene may encode any protein, including those described

5   above.

Design and assembly of the DNA constructs

Constructs may be designed in accordance with the principles, illustrative examples and materials and methods disclosed in the patent documents and scientific literature cited herein, with

10   modifications and further exemplification as described. Components of the constructs can be prepared in conventional ways, where the coding sequences and regulatory regions may be isolated, as appropriate, ligated, cloned in an appropriate cloning host, analyzed by restriction or sequencing, or other convenient means. Particularly, using PCR, individual fragments including all or portions of a functional unit may be isolated, where one or more mutations may be introduced using "primer repair",

15   ligation, in vitro mutagenesis, etc. as appropriate. In the case of DNA constructs encoding fusion proteins, DNA sequences encoding individual domains and sub-domains are joined such that they constitute a single open reading frame encoding a fusion protein capable of being translated in cells or cell lysates into a single polypeptide harboring all component domains. The DNA construct encoding the fusion protein may then be placed into a vector for transducing host cells and permitting the

20   expression of the protein. For biochemical analysis of the encoded chimera, it may be desirable to construct plasmids that direct the expression of the protein in bacteria or in reticulocyte-lysate systems. For use in the production of proteins in mammalian cells, the protein-encoding sequence is introduced into an expression vector that directs expression in these cells. Expression vectors suitable for such uses are well known in the art. Various sorts of such vectors are commercially

25   available.

Promoters

The fusion proteins described herein may be used in combination with any promoter that will direct their expression in mammalian cells. The promoter may be a strong promoter, such as the

30   human CMV promoter , or a weaker promoter, such as a promoter for an endogenous human gene. Other promoters which may be used include, but are not limited to, the Rous Sarcoma Virus (RSV) promoter, the retroviral LTR from Murine Moloney Leukemia Virus (MMLV), the muscle creatine kinase (MCK) enhancer, the SV40 promoter, and the CMV enhancer from the major immediate early gene. Genbank accession numbers for the above promoters are given in the table below.

Promoter	Genbank Accession Number
CMV	AF067197

17

RSV	M83237
MMLV LTR	M77239
SV40	U47120
CMV enhancer for MIE gene	K03104
MCK enhancer	X67536

In many cases, the selection of promoter will depend upon the configuration of the fusion protein used in a particular application. Thus, if the practitioner desired the CAD-containing fusion protein to be expressed at high levels, a stronger promoter, such as CMV, would be used. Alternatively, for tissue specific expression, a tissue specific promoter like the MCK enhancer (for expression in muscle) would be selected.

Introduction of Constructs into Cells

This invention is particularly useful for the engineering of animal cells and in applications involving the use of such engineered animal cells. The animal cells may be, among others, insect, worm or mammalian cells. While various mammalian cells may be used, including, by way of example, equine, bovine, ovine, canine, feline, murine, and non-human primate cells, human and mouse cells are of particular interest. Across the various species, various types of cells may be used, such as hematopoietic, neural, glial, mesenchymal, cutaneous, mucosal, stromal, muscle (including smooth muscle cells), spleen, reticuloendothelial, epithelial, endothelial, hepatic, kidney, gastrointestinal, pulmonary, fibroblast, and other cell types. Of particular interest are muscle cells (including skeletal, cardiac and other muscle cells), cells of the central and peripheral nervous systems, and hematopoietic cells, which may include any of the nucleated cells which may be involved with the erythroid, lymphoid or myelomonocytic lineages, as well as myoblasts and fibroblasts. Also of interest are stem and progenitor cells, such as hematopoietic, neural, stromal, muscle, hepatic, pulmonary, gastrointestinal and mesenchymal stem cells

The cells may be autologous cells, syngeneic cells, allogeneic cells and even in some cases, xenogeneic cells with respect to an intended host organism. The cells may be modified by changing the major histocompatibility complex ("MHC") profile, by inactivating β2-microglobulin to prevent the formation of functional Class I MHC molecules, inactivation of Class II molecules, providing for expression of one or more MHC molecules, enhancing or inactivating cytotoxic capabilities by enhancing or inhibiting the expression of genes associated with the cytotoxic activity, and the like.

In some instances specific clones or oligoclonal cells may be of interest, where the cells have a particular specificity, such as T cells and B cells having a specific antigen specificity or homing target site specificity.

Constructs encoding the fusion proteins and comprising target genes of this invention can be introduced into the cells as one or more nucleic acid molecules or constructs, in many cases in association with one or more markers to allow for selection of host cells which contain the construct(s). The constructs can be prepared in conventional ways, where the coding sequences and regulatory

5   regions may be isolated, as appropriate, ligated, cloned in an appropriate cloning host, analyzed by restriction or sequencing, or other convenient means. Particularly, using PCR, individual fragments including all or portions of a functional domain may be isolated, where one or more mutations may be introduced using "primer repair", ligation, *in vitro* mutagenesis, *etc.* as appropriate.

The construct(s) once completed and demonstrated to have the appropriate sequences may

10   then be introduced into a host cell by any convenient means. The constructs may be incorporated into vectors capable of episomal replication (e.g. BPV or EBV vectors) or into vectors designed for integration into the host cells' chromosomes. The constructs may be integrated and packaged into non-replicating, defective viral genomes like Adenovirus, Adeno-associated virus (AAV), or Herpes simplex virus (HSV) or others, including retroviral vectors, for infection or transduction into cells.

15   Alternatively, the construct may be introduced by protoplast fusion, electroporation, biolistics, calcium phosphate transfection, lipofection, microinjection of DNA or the like. The host cells will in some cases be grown and expanded in culture before introduction of the construct(s), followed by the appropriate treatment for introduction of the construct(s) and integration of the construct(s). The cells may then be expanded and/or screened by virtue of a marker present in the constructs. Various markers which

20   may be used successfully include *hprt*, neomycin resistance, thymidine kinase, hygromycin resistance, *etc.*, and various cell-surface markers such as Tac, CD8, CD3, Thy1 and the NGF receptor.

In some instances, one may have a target site for homologous recombination, where it is desired that a construct be integrated at a particular locus. For example, one can delete and/or replace an endogenous gene (at the same locus or elsewhere) with a recombinant target construct of this

25   invention. For homologous recombination, one may generally use either Ω or O-vectors. See, for example, Thomas and Capecchi, *Cell* (1987) 51, 503-512; Mansour, *et al., Nature* (1988) 336, 348-352; and Joyner, *et al., Nature* (1989) 338, 153-156.

The constructs may be introduced as a single DNA molecule encoding all of the genes, or different DNA molecules having one or more genes. The constructs may be introduced

30   simultaneously or consecutively, each with the same or different markers.

Vectors containing useful elements such as bacterial or yeast origins of replication, selectable and/or amplifiable markers, promoter/enhancer elements for expression in prokaryotes or eukaryotes, and mammalian expression control elements, etc. which may be used to prepare stocks of construct DNAs and for carrying out transfections are well known in the art, and many are commercially available.

35

Introduction of Constructs into Animals

Any means for the introduction of genetically engineered cells or heterologous DNA into animals, preferably mammals, human or non-human, may be adapted to the practice of this invention

for the delivery of the various DNA constructs into the intended recipient. For the purpose of this discussion, the various DNA constructs described herein may together be referred to as the transgene.

5   *by ex vivo genetic engineering*

Cells which have been transduced ex vivo or in vitro with the DNA constructs may be grown in culture under selective conditions and cells which are selected as having the desired construct(s) may then be expanded and further analyzed, using, for example, the polymerase chain reaction for determining the presence of the construct in the host cells and/or assays for the production of the

10   desired gene product(s). After being transduced with the heterologous genetic constructs, the modified host cells may be identified, selected, grown, characterized, etc. as desired, and then may be used as planned, e.g. grown in culture or introduced into a host organism.

Depending upon the nature of the cells, the cells may be introduced into a host organism, e.g. a mammal, in a wide variety of ways, generally by injection or implantation into the desired tissue or

15   compartment, or a tissue or compartment permitting migration of the cells to their intended destination. Illustrative sites for injection or implantation include the vascular system, bone marrow, muscle, liver, cranium or spinal cord, peritoneum, and skin. Hematopoietic cells, for example, may be administered by injection into the vascular system, there being usually at least about 10^4 cells and generally not more than about 10^{10} cells. The number of cells which are employed will depend upon

20   the circumstances, the purpose for the introduction, the lifetime of the cells, the protocol to be used, for example, the number of administrations, the ability of the cells to multiply, the stability of the therapeutic agent, the physiologic need for the therapeutic agent, and the like. Generally, for myoblasts or fibroblasts for example, the number of cells will be at least about 10^4 and not more than about 10^9 and may be applied as a dispersion, generally being injected at or near the site of interest.

25   The cells will usually be in a physiologically-acceptable medium.

Cells engineered in accordance with this invention may also be encapsulated, e.g. using conventional biocompatible materials and methods, prior to implantation into the host organism or patient for the production of a therapeutic protein. See e.g. Hguyen et al, Tissue Implant Systems and Methods for Sustaining viable High Cell Densities within a Host, US Patent No. 5,314,471 (Baxter

30   International, Inc.); Uludag and Sefton, 1993, J Biomed. Mater. Res. 27(10):1213-24 (HepG2 cells/hydroxyethyl methacrylate-methyl methacrylate membranes); Chang et al, 1993, Hum Gene Ther 4(4):433-40 (mouse Ltk- cells expressing hGH/immunoprotective perm-selective alginate microcapsules; Reddy et al, 1993, J Infect Dis 168(4):1082-3 (alginate); Tai and Sun, 1993, FASEB J 7(11):1061-9 (mouse fibroblasts expressing hGH/alginate-poly-L-lysine-alginate membrane); Ao et al,

35   1995, Transplantation Proc. 27(6):3349, 3350 (alginate); Rajotte et al, 1995, Transplantation Proc. 27(6):3389 (alginate); Lakey et al, 1995, Transplantation Proc. 27(6):3266 (alginate); Korbutt et al, 1995, Transplantation Proc. 27(6):3212 (alginate); Dorian et al, US Patent No. 5,429,821 (alginate);

Emerich et al, 1993, Exp Neurol 122(1):37-47 (polymer-encapsulated PC12 cells); Sagen et al, 1993, J Neurosci 13(6):2415-23 (bovine chromaffin cells encapsulated in semipermeable polymer membrane and implanted into rat spinal subarachnoid space); Aebischer et al, 1994, Exp Neurol 126(2):151-8 (polymer-encapsulated rat PC12 cells implanted into monkeys; see also Aebischer, WO

5   92/19595); Savelkoul et al, 1994, J Immunol Methods 170(2):185-96 (encapsulated hybridomas producing antibodies; encapsulated transfected cell lines expressing various cytokines); Winn et al, 1994, PNAS USA 91(6):2324-8 (engineered BHK cells expressing human nerve growth factor encapsulated in an immunoisolation polymeric device and transplanted into rats); Emerich et al, 1994, Prog Neuropsychopharmacol Biol Psychiatry 18(5):935-46 (polymer-encapsulated PC12 cells

10   implanted into rats); Kordower et al, 1994, PNAS USA 91(23):10898-902 (polymer-encapsulated engineered BHK cells expressing hNGF implanted into monkeys) and Butler et al WO 95/04521 (encapsulated device). The cells may then be introduced in encapsulated form into an animal host, preferably a mammal and more preferably a human subject in need thereof. Preferably the encapsulating material is semipermeable, permitting release into the host of taget proteins produced

15   by the encapsulated cells. In many embodiments the semipermeable encapsulation renders the encapsulated cells immunologically isolated from the host organism in which the encapsulated cells are introduced. In those embodiments the cells to be encapsulated may express one or more fusion proteins containing component domains derived from proteins of the host species and/or from viral proteins or proteins from species other than the host species. The cells may be derived from one or

20   more individuals other than the recipient and may be derived from a species other than that of the recipient organism or patient.

by in vivo genetic engineering

Instead of ex vivo modification of the cells, in many situations one may wish to modify cells *in*

25   *vivo*. A variety of techniques have been developed for genetic engineering of target tissue and cells *in vivo*, including viral and non-viral systems.

In one approach, the DNA constructs are delivered to cells by transfection, i.e., by delivery to cells of "naked DNA", lipid-complexed or liposome-formulated DNA, or otherwise formulated DNA. Prior to formulation of DNA, e.g., with lipid, or as in other approaches, prior to incorporation in a final

30   expression vector, a plasmid containing a transgene bearing the desired DNA constructs may first be experimentally optimized for expression (e.g., inclusion of an intron in the 5' untranslated region and elimination of unnecessary sequences (Felgner, et al., Ann NY Acad Sci 126-139, 1995). Formulation of DNA, e.g. with various lipid or liposome materials, may then be effected using known methods and materials and delivered to the recipient mammal. See, e.g., Canonico et al, Am J Respir Cell Mol Biol

35   10:24-29, 1994 (in vivo transfer of an aerosolized recombinant human alpha1-antitrypsin gene complexed to cationic liposomes to the lungs of rabbits); Tsan et al, Am J Physiol 268 (Lung Cell Mol Physiol 12): L1052-L1056, 1995 (transfer of genes to rat lungs via tracheal insufflation of plasmid DNA alone or complexed with cationic liposomes); Alton et al., Nat Genet. 5:135-142, 1993 (gene transfer

to mouse airways by nebulized delivery of cDNA-liposome complexes). In either case, delivery of vectors or naked or formulated DNA can be carried out by instillation via bronchoscopy, after transfer of viral particles to Ringer's, phosphate buffered saline, or other similar vehicle, or by nebulization.

Viral systems include those based on viruses such as adenovirus, adeno-associated virus,
5   hybrid adeno-AAV, lentivirus and retroviruses, which allow for transduction by infection, and in some cases, integration of the virus or transgene into the host genome. See, for example, Dubensky et al. (1984) Proc. Natl. Acad. Sci. USA 81, 7529-7533; Kaneda et al., (1989) Science 243,375-378; Hiebert et al. (1989) Proc. Natl. Acad. Sci. USA 86, 3594-3598; Hatzoglu et al. (1990) J. Biol. Chem. 265, 17285-17293 and Ferry, et al. (1991) Proc. Natl. Acad. Sci. USA 88, 8377-8381. The virus may
10   be administered by injection (e.g. intravascularly or intramuscularly), inhalation, or other parenteral mode. Non-viral delivery methods such as administration of the DNA via complexes with liposomes or by injection, catheter or biolistics may also be used. See e.g. WO 96/41865, PCT/US97/22454 and USSN 60/084819, for example, for additional guidance on formulation and delivery of recombinant nucleic acids to cells and to organisms.

15   By employing an attenuated or modified retrovirus carrying a target transcriptional initiation region, if desired, one can activate the virus using one of the subject transcription factor constructs, so that the virus may be produced and transduce adjacent cells.

The use of recombinant viruses to deliver the nucleic acid constructs are of particular interest. The transgene(s) may be incorporated into any of a variety of viruses useful in gene therapy.
20   In clinical settings, the gene delivery systems (i.e., the recombinant nucleic acids in vectors, virus, lipid formulation or other form) can be introduced into a patient, e.g., by any of a number of known methods. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, inhalation, etc. In some systems, the means of delivery provides for specific or selective transduction of the construct into desired target cells. This
25   can be achieved by regional or local administration (see U.S. Patent 5,328,470) or by stereotactic injection, e.g. Chen et al., (1994) PNAS USA 91: 3054-3057 or by determinants of the delivery means. For instance, some viral systems have a tissue or cell-type specificity for infection. In some systems cell-type or tissue-type expression is achieved by the use of cell-type or tissue-specific expression control elements controlling expression of the gene.
30   In preferred embodiments of the invention, the subject expression constructs are derived by incorporation of the genetic construct(s) of interest into viral delivery systems including a recombinant retrovirus, adenovirus, adeno-associated virus (AAV), hybrid adenovirus/AAV, herpes virus or lentivirus (although other applications may be carried out using recombinant bacterial or eukaryotic plasmids). While various viral vectors may be used in the practice of this invention, AAV- and
35   adenovirus-based approaches are of particular interest for the transfer of exogenous genes in vivo, particularly into humans and other mammals. The following additional guidance on the choice and use of viral vectors may be helpful to the practitioner, especially with respect to applications involving whole animals (including both human gene therapy and the development and use of animal model

22

systems), whether ex vivo or in vivo.

Viral Vect rs:

5   **Adenoviral vectors**

A viral gene delivery system useful in the present invention utilizes adenovirus-derived vectors. Knowledge of the genetic organization of adenovirus, a 36 kb, linear and double-stranded DNA virus, allows substitution of a large piece of adenoviral DNA with foreign sequences up to 8 kb. In contrast to retrovirus, the infection of adenoviral DNA into host cells does not result in chromosomal

10   integration because adenoviral DNA can replicate in an episomal manner without potential genotoxicity. Also, adenoviruses are structurally stable, and no genome rearrangement has been detected after extensive amplification. Adenovirus can infect virtually all epithelial cells regardless of their cell cycle stage. So far, adenoviral infection appears to be linked only to mild disease such as acute respiratory disease in the human.

15   Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized genome, ease of manipulation, high titer, wide target-cell range, and high infectivity. Both ends of the viral genome contain 100-200 base pair (bp) inverted terminal repeats (ITR), which are cis elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome contain different transcription domains that are divided by the onset of viral DNA replication. The E1

20   region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication. These proteins are involved in DNA replication, late gene expression, and host cell shut off (Renan (1990) Radiotherap. Oncol. 19:197). The products of the late genes, including the majority of the viral capsid proteins, are expressed only after significant

25   processing of a single primary transcript issued by the major late promoter (MLP). The MLP (located at 16.8 m.u.) is particularly efficient during the late phase of infection, and all the mRNAs issued from this promoter possess a 5' tripartite leader (TL) sequence which makes them preferred mRNAs for translation.

The genome of an adenovirus can be manipulated such that it encodes a gene product of interest, but is inactivated in terms of its ability to replicate in a normal lytic viral life cycle (see, for

30   example, Berkner et al., (1988) BioTechniques 6:616; Rosenfeld et al., (1991) Science 252:431-434; and Rosenfeld et al., (1992) Cell 68:143-155). Suitable adenoviral vectors derived from the adenovirus strain Ad type 5 dl324 or other strains of adenovirus (e.g., Ad2, Ad3, Ad7 etc.) are well known to those skilled in the art. Recombinant adenoviruses can be advantageous in certain

35   circumstances in that they are not capable of infecting nondividing cells and can be used to infect a wide variety of cell types, including airway epithelium (Rosenfeld et al., (1992) cited *supra*), endothelial cells (Lemarchand et al., (1992) PNAS USA 89:6482-6486), hepatocytes (Herz and Gerard, (1993) PNAS USA 90:2812-2816) and muscle cells (Quantin et al., (1992) PNAS USA 89:2581-2584).

Adenovirus vectors have also been used in vaccine development (Grunhaus and Horwitz (1992) Seminar in Virology 3:237; Graham and Prevec (1992) Biotechnology 20:363). Experiments in administering recombinant adenovirus to different tissues include trachea instillation (Rosenfeld et al. (1991) ; Rosenfeld et al. (1992) Cell 68:143), muscle injection (Ragot et al. (1993) Nature 361:647),

5   peripheral intravenous injection (Herz and Gerard (1993) Proc. Natl. Acad. Sci. U.S.A. 90:2812), and stereotactic inoculation into the brain (Le Gal La Salle et al. (1993) Science 254:988).

Furthermore, the virus particle is relatively stable and amenable to purification and concentration, and as above, can be modified so as to affect the spectrum of infectivity. Additionally, adenovirus is easy to grow and manipulate and exhibits broad host range in vitro and in vivo. This

10   group of viruses can be obtained in high titers, e.g., 10^9 - 10^{11} plaque-forming unit (PFU)/ml, and they are highly infective. The life cycle of adenovirus does not require integration into the host cell genome. The foreign genes delivered by adenovirus vectors are episomal, and therefore, have low genotoxicity to host cells. No side effects have been reported in studies of vaccination with wild-type adenovirus (Couch et al., 1963; Top et al., 1971), demonstrating their safety and therapeutic potential

15   as in vivo gene transfer vectors. Moreover, the carrying capacity of the adenoviral genome for foreign DNA is large (up to 8 kilobases) relative to other gene delivery vectors (Berkner et al., supra; Haj-Ahmand and Graham (1986) J. Virol. 57:267). Most replication-defective adenoviral vectors currently in use and therefore favored by the present invention are deleted for all or parts of the viral E1 and E3 genes but retain as much as 80% of the adenoviral genetic material (see, e.g., Jones et al., (1979) Cell

20   16:683; Berkner et al., supra; and Graham et al., in Methods in Molecular Biology, E.J. Murray, Ed. (Humana, Clifton, NJ, 1991) vol. 7. pp. 109-127). Expression of the inserted gene can be under control of, for example, the E1A promoter, the major late promoter (MLP) and associated leader sequences, the viral E3 promoter, or exogenously added promoter sequences.

Other than the requirement that the adenovirus vector be replication defective, or at least

25   conditionally defective, the nature of the adenovirus vector is not believed to be crucial to the successful practice of the invention. The adenovirus may be of any of the 42 different known serotypes or subgroups A-F. Adenovirus type 5 of subgroup C is the preferred starting material in order to obtain the conditional replication-defective adenovirus vector for use in the method of the present invention. This is because Adenovirus type 5 is a human adenovirus about which a great

30   deal of biochemical and genetic information is known, and it has historically been used for most constructions employing adenovirus as a vector. As stated above, the typical vector according to the present invention is replication defective and will not have an adenovirus E1 region. Thus, it will be most convenient to introduce the nucleic acid of interest at the position from which the E1 coding sequences have been removed. However, the position of insertion of the nucleic acid of interest in a

35   region within the adenovirus sequences is not critical to the present invention. For example, the nucleic acid of interest may also be inserted in lieu of the deleted E3 region in E3 replacement vectors as described previously by Karlsson et. al. (1986) or in the E4 region where a helper cell line or helper virus complements the E4 defect.

A preferred helper cell line is 293 (ATCC Accession No. CRL1573). This helper cell line, also termed a "packaging cell line" was developed by Frank Graham (Graham et al. (1987) J. Gen. Virol. 36:59-72 and Graham (1977) J.General Virology 68:937-940) and provides E1A and E1B in trans. However, helper cell lines may also be derived from human cells such as human embryonic kidney

5   cells, muscle cells, hematopoietic cells or other human embryonic mesenchymal or epithelial cells. Alternatively, the helper cells may be derived from the cells of other mammalian species that are permissive for human adenovirus. Such cells include, e.g., Vero cells or other monkey embryonic mesenchymal or epithelial cells.

Various adenovirus vectors have been shown to be of use in the transfer of genes to

10   mammals, including humans. Replication-deficient adenovirus vectors have been used to express marker proteins and CFTR in the pulmonary epithelium. Because of their ability to efficiently infect dividing cells, their tropism for the lung, and the relative ease of generation of high titer stocks, adenoviral vectors have been the subject of much research in the last few years, and various vectors have been used to deliver genes to the lungs of human subjects (Zabner et al., Cell 75:207-216,

15   1993; Crystal, et al., Nat Genet. 8:42-51, 1994; Boucher, et al., Hum Gene Ther 5:615-639, 1994). The first generation E1a deleted adenovirus vectors have been improved upon with a second generation that includes a temperature-sensitive E2a viral protein, designed to express less viral protein and thereby make the virally infected cell less of a target for the immune system (Goldman et al., Human Gene Therapy 6:839-851,1995). More recently, a viral vector deleted of all viral open

20   reading frames has been reported (Fisher et al., Virology 217:11-22, 1996). Moreover, it has been shown that expression of viral IL-10 inhibits the immune response to adenoviral antigen (Qin et al., Human Gene Therapy 8:1365-1374, 1997).

Adenoviruses can also be cell type specific, i.e., infect only restricted types of cells and/or express a transgene only in restricted types of cells. For example, the viruses comprise a gene under

25   the transcriptional control of a transcription initiation region specifically regulated by target host cells, as described e.g., in U.S. Patent No. 5,698,443, by Henderson and Schuur, issued December 16, 1997. Thus, replication competent adenoviruses can be restricted to certain cells by, e.g., inserting a cell specific response element to regulate a synthesis of a protein necessary for replication, e.g., E1A or E1B.

30   DNA sequences of a number of adenovirus types are available from Genbank. For example, human adenovirus type 5 has GenBank Accession No.M73260. The adenovirus DNA sequences may be obtained from any of the 42 human adenovirus types currently identified. Various adenovirus strains are available from the American Type Culture Collection, Rockville, Maryland, or by request from a number of commercial and academic sources. A transgene as described herein may be incorporated

35   into any adenoviral vector and delivery protocol, by the same methods (restriction digest, linker ligation or filling in of ends, and ligation) used to insert the CFTR or other genes into the vectors.

Adenovirus producer cell lines can include one or more of the adenoviral genes E1, E2a, and E4 DNA sequence, for packaging adenovirus vectors in which one or more of these genes have

be n mutated or deleted are described, e.g., in PCT/US95/15947 (WO 96/18418) by Kadan et al.; PCT/US95/07341 (WO 95/346671) by Kovesdi et al.; PCT/FR94/00624 (WO94/28152) by Imler et al.;PCT/FR94/00851 (WO 95/02697) by Perrocaudet et al., PCT/US95/14793 (WO96/14061) by Wang et al.

AAV Vectors

Another viral vector system useful for delivery of DNA is the adeno-associated virus (AAV). Adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle. (For a review, see Muzyczka et al., Curr. Topics in Micro. and Immunol. (1992) 158:97-129).

AAV has not been associated with the cause of any disease. AAV is not a transforming or oncogenic virus. AAV integration into chromosomes of human cell lines does not cause any significant alteration in the growth properties or morphological characteristics of the cells. These properties of AAV also recommend it as a potentially useful human gene therapy vector.

AAV is also one of the few viruses that may integrate its DNA into non-dividing cells, e.g., pulmonary epithelial cells or muscle cells, and exhibits a high frequency of stable integration (see for example Flotte et al., (1992) Am. J. Respir. Cell. Mol. Biol. 7:349-356; Samulski et al., (1989) J. Virol. 63:3822-3828; and McLaughlin et al., (1989) J. Virol. 62:1963-1973). Vectors containing as little as 300 base pairs of AAV can be packaged and can integrate. Space for exogenous DNA is limited to about 4.5 kb. An AAV vector such as that described in Tratschin et al., (1985) Mol. Cell. Biol. 5:3251-3260 can be used to introduce DNA into cells. A variety of nucleic acids have been introduced into different cell types using AAV vectors (see for example Hermonat et al., (1984) PNAS USA 81:6466-6470; Tratschin et al., (1985) Mol. Cell. Biol. 4:2072-2081; Wondisford et al., (1988) Mol. Endocrinol. 2:32-39; Tratschin et al., (1984) J. Virol. 51:611-619; and Flotte et al., (1993) J. Biol. Chem. 268:3781-3790).

The AAV-based expression vector to be used typically includes the 145 nucleotide AAV inverted terminal repeats (ITRs) flanking a restriction site that can be used for subcloning of the transgene, either directly using the restriction site available, or by excision of the transgene with restriction enzymes followed by blunting of the ends, ligation of appropriate DNA linkers, restriction digestion, and ligation into the site between the ITRs. The capacity of AAV vectors is about 4.4 kb. The following proteins have been expressed using various AAV-based vectors, and a variety of promoter/enhancers: neomycin phosphotransferase, chloramphenicol acetyl transferase, Fanconi's anemia gene, cystic fibrosis transmembrane conductance regulator, and granulocyte macrophage colony-stimulating factor (Kotin, R.M., Human Gene Therapy 5:793-801, 1994, Table I). A transgene incorporating the various DNA constructs of this invention can similarly be included in an AAV-based vector. As an alternative to inclusion of a constitutive promoter such as CMV to drive expression of the recombinant DNA encoding the fusion protein(s), e.g. fusion proteins comprising an activation domain or DNA-binding domain, an AAV promoter can be used (ITR itself or AAV p5 (Flotte, et al. J.

Biol.Chem. 268:3781-3790, 1993)).

Such a vector can be packaged into AAV virions by reported methods. For example, a human cell line such as 293 can be co-transfected with the AAV-based expression vector and another plasmid containing open reading frames encoding AAV rep and cap (which are obligatory for

5   replication and packaging of the recombinant viral construct) under the control of endogenous AAV promoters or a heterologous promoter. In the absence of helper virus, the rep proteins Rep68 and Rep78 prevent accumulation of the replicative form, but upon superinfection with adenovirus or herpes virus, these proteins permit replication from the ITRs (present only in the construct containing the transgene) and expression of the viral capsid proteins. This system results in packaging of the

10   transgene DNA into AAV virions (Carter, B.J., Current Opinion in Biotechnology 3:533-539, 1992; Kotin, R.M, Human Gene Therapy 5:793-801, 1994)). Typically, three days after transfection, recombinant AAV is harvested from the cells along with adenovirus and the contaminating adenovirus is then inactivated by heat treatment.

Methods to improve the titer of AAV can also be used to express the transgene in an AAV

15   virion. Such strategies include, but are not limited to: stable expression of the ITR-flanked transgene in a cell line followed by transfection with a second plasmid to direct viral packaging; use of a cell line that expresses AAV proteins inducibly, such as temperature-sensitive inducible expression or pharmacologically inducible expression. Alternatively, a cell can be transformed with a first AAV vector including a 5' ITR, a 3' ITR flanking a heterologous gene, and a second AAV vector which includes an

20   inducible origin of replication, e.g., SV40 origin of replication, which is capable of being induced by an agent, such as the SV40 T antigen and which includes DNA sequences encoding the AAV rep and cap proteins. Upon induction by an agent, the second AAV vector may replicate to a high copy number, and thereby increased numbers of infectious AAV particles may be generated (see, e.g, U.S. Patent No. 5,693,531 by Chiorini et al., issued December 2, 1997. In yet another method for

25   producing large amounts of recombinant AAV, a plasmid is used which incorporate the Epstein Barr Nuclear Antigen (EBNA) gene , the latent origin of replication of Epstein Barr virus (oriP) and an AAV genome. These plasmids are maintained as a multicopy extra-chromosomal elements in cells, such as in 293 cells. Upon addition of wild-type helper functions, these cells will produce high amounts of recombinant AAV (U.S. Patent 5,691,176 by Lebkowski et al., issued Nov. 25, 1997). In another

30   system, an AAV packaging plasmid is provided that allows expression of the rep gene, wherein the p5 promoter, which normally controls rep expression, is replaced with a heterologous promoter (U.S. Patent 5,658,776, by Flotte et al., issued Aug. 19, 1997). Additionally, one may increase the efficiency of AAV transduction by treating the cells with an agent that facilitates the conversion of the single stranded form to the double stranded form, as described in Wilson et al., WO96/39530.

35   AAV stocks can be produced as described in Hermonat and Muzyczka (1984) PNAS 81:6466, modified by using the pAAV/Ad described by Samulski et al. (1989) J. Virol. 63:3822. Concentration and purification of the virus can be achieved by reported methods such as banding in cesium chloride gradients, as was used for the initial report of AAV vector expression in vivo (Flotte, et

al. J.Biol. Chem. 268:3781-3790, 1993) or chromatographic purification, as described in O'Riordan et al., WO97/08298.

Methods for *in vitro* packaging AAV vectors are also available and have the advantage that there is no size limitation of the DNA packaged into the particles (see, U.S. Patent No. 5,688,676, by
5   Zhou et al., issued Nov. 18, 1997). This procedure involves the preparation of cell free packaging extracts.

For additional detailed guidance on AAV technology which may be useful in the practice of the subject invention, including methods and materials for the incorporation of a transgene, the propagation and purification of the recombinant AAV vector containing the transgene, and its use in
10   transfecting cells and mammals, see e.g. Carter et al, US Patent No. 4,797,368 (10 Jan 1989); Muzyczka et al, US Patent No. 5,139,941 (18 Aug 1992); Lebkowski et al, US Patent No. 5,173,414 (22 Dec 1992); Srivastava, US Patent No. 5,252,479 (12 Oct 1993); Lebkowski et al, US Patent No. 5,354,678 (11 Oct 1994); Shenk et al, US Patent No. 5,436,146(25 July 1995); Chatterjee et al, US Patent No. 5,454,935 (12 Dec 1995), Carter et al WO 93/24641 (published 9 Dec 1993), and
15   Natsoulis, U.S. Patent No. 5,622,856 (April 22, 1997). Further information regarding AAVs and the adenovirus or herpes helper functions required can be found in the following articles.Berns and Bohensky (1987), "Adeno-Associated Viruses: An Update", Advanced in Virus Research, Academic Press, 33:243-306. The genome of AAV is described in Laughlin et al. (1983) "Cloning of infectious adeno-associated virus genomes in bacterial plasmids", Gene, 23: 65-73. Expression of AAV is
20   described in Beaton et al. (1989) "Expression from the Adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein", J. Virol., 63:4450-4454. Construction of rAAV is described in a number of publications: Tratschin et al. (1984) "Adeno-associated virus vector for high frequency integration, expression and rescue of genes in mammalian cells", Mol. Cell. Biol., 4:2072-2081; Hermonat and Muzyczka (1984) "Use of adeno-associated virus as a mammalian DNA
25   cloning vector: Transduction of neomycin resistance into mammalian tissue culture cells", Proc. Natl. Acad. Sci. USA, 81:6466-6470; McLaughlin et al. (1988) "Adeno-associated virus general transduction vectors: Analysis of Proviral Structures", J. Virol., 62:1963-1973; and Samulski et al. (1989) "Helper-free stocks of recombinant adeno-associated viruses: normal integration doesenquire viral gene expression", J. Virol., 63:3822-3828. Cell lines that can be transformed by rAAV are those
30   described in Lebkowski et al. (1988) "Adeno-associated virus: a vector system for efficient introduction and integration of DNA into a variety of mammalian cell types", Mol. Cell. Biol., 8:3988-3996. "Producer" or "packaging" cell lines used in manufacturing recombinant retroviruses are described in Dougherty et al. (1989) J. Virol., 63:3209-3212; and Markowitz et al. (1988) J. Virol., 62:1120-1124.

35

Hybrid Adenovirus-AAV Vectors

Hybrid Adenovirus-AAV vectors represented by an adenovirus capsid containing a nucleic acid comprising a portion of an adenovirus, and 5' and 3' ITR sequences from an AAV which flank a

selected transgene under the control of a promoter. See e.g. Wilson et al, International Patent
Application Publication No. WO 96/13598. This hybrid vector is characterized by high titer transgene
delivery to a host cell and the ability to stably integrate the transgene into the host cell chromosome in
the presence of the rep gene. This virus is capable of infecting virtually all cell types (conferred by its

5   adenovirus sequences) and stable long term transgene integration into the host cell genome
(conferred by its AAV sequences).

The adenovirus nucleic acid sequences employed in the this vector can range from a
minimum sequence amount, which requires the use of a helper virus to produce the hybrid virus
particle, to only selected deletions of adenovirus genes, which deleted gene products can be

10   supplied in the hybrid viral process by a packaging cell. For example, a hybrid virus can comprise the 5'
and 3' inverted terminal repeat (ITR) sequences of an adenovirus (which function as origins of
replication). The left terminal sequence (5') sequence of the Ad5 genome that can be used spans bp
1 to about 360 of the conventional adenovirus genome (also referred to as map units 0-1) and
includes the 5' ITR and the packaging/enhancer domain. The 3' adenovirus sequences of the hybrid

15   virus include the right terminal 3' ITR sequence which is about 580 nucleotides (about bp 35,353- end
of the adenovirus, referred to as about map units 98.4-100.

The AAV sequences useful in the hybrid vector are viral sequences from which the rep and
cap polypeptide encoding sequences are deleted and are usually the cis acting 5' and 3' ITR
sequences. Thus, the AAV ITR sequences are flanked by the selected adenovirus sequences and

20   the AAV ITR sequences themselves flank a selected transgene. The preparation of the hybrid vector
is further described in detail in published PCT application entitled "Hybrid Adenovirus-AAV Virus and
Method of Use Thereof", WO 96/13598 by Wilson et al.

For additional detailed guidance on adenovirus and hybrid adenovirus-AAV technology which
may be useful in the practice of the subject invention, including methods and materials for the

25   incorporation of a transgene, the propagation and purification of recombinant virus containing the
transgene, and its use in transfecting cells and mammals, see also Wilson et al, WO 94/28938, WO
96/13597 and WO 96/26285, and references cited therein.

Retrovirus s

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells by a process of reverse-transcription

5   (Coffin (1990) Retroviridae and their Replication" In Fields, Knipe ed. Virology. New York: Raven Press). The resulting DNA then stably integrates into cellular chromosomes as a provirus and directs synthesis of viral proteins. The integration results in the retention of the viral gene sequences in the recipient cell and its descendants. The retroviral genome contains three genes, gag, pol, and env that code for capsidal proteins, polymerase enzyme, and envelope components, respectively. A

10   sequence found upstream from the gag gene, termed psi , functions as a signal for packaging of the genome into virions. Two long terminal repeat (LTR) sequences are present at the 5' and 3' ends of the viral genome. These contain strong promoter and enhancer sequences and are also required for integration in the host cell genome (Coffin (1990), supra).

In order to construct a retroviral vector, a nucleic acid of interest is inserted into the viral

15   genome in the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the gag, pol, and env genes but without the LTR and psi components is constructed (Mann et al. (1983) Cell 33:153). When a recombinant plasmid containing a human cDNA, together with the retroviral LTR and psi sequences is introduced into this cell line (by calcium phosphate precipitation for example), the psi sequence allows the RNA transcript

20   of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas and Rubenstein (1988) "Retroviral Vectors", In: Rodriguez and Denhardt ed. Vectors: A Survey of Molecular Cloning Vectors and their Uses. Stoneham:Butterworth; Temin, (1986) "Retrovirus Vectors for Gene Transfer: Efficient Integration into and Expression of Exogenous DNA in Vertebrate Cell Genome", In: Kucherlapati ed. Gene Transfer. New York: Plenum Press; Mann et al.,

25   1983, supra). The media containing the recombinant retroviruses is then collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However, integration and stable expression require the division of host cells (Paskind et al. (1975) Virology 67:242).

A major prerequisite for the use of retroviruses is to ensure the safety of their use, particularly

30   with regard to the possibility of the spread of wild-type virus in the cell population. The development of specialized cell lines (termed "packaging cells") which produce only replication-defective retroviruses has increased the utility of retroviruses for gene therapy, and defective retroviruses are well characterized for use in gene transfer for gene therapy purposes (for a review see Miller, A.D. (1990) Blood 76:271). Thus, recombinant retrovirus can be constructed in which part of the retroviral

35   coding sequence (gag, pol, env) has been replaced by nucleic acid encoding a fusion protein of the present invention, rendering the retrovirus replication defective. The replication defective retrovirus is then packaged into virions which can be used to infect a target cell through the use of a helper virus by standard techniques. Protocols for producing recombinant retroviruses and for infecting cells in vitro

or in vivo with such virus s can be found in Current Protocols in Molecular Biology, Ausubel, F.M. et al., (eds.) Greene Publishing Associates, (1989), Sections 9.10-9.14 and other standard laboratory manuals. Examples of suitable retroviruses include pLJ, pZIP, pWE and pEM which are well known to those skilled in the art. A preferred retroviral vector is a pSR MSVtkNeo (Muller et al. (1991) Mol. Cell

5   Biol. 11:1785 and pSR MSV(XbaI) (Sawyers et al. (1995) J. Exp. Med. 181:307) and derivatives thereof. For example, the unique BamHI sites in both of these vectors can be removed by digesting the vectors with BamHI, filling in with Klenow and religating to produce pSMTN2 and pSMTX2, respectively, as described in PCT/US96/09948 by Clackson et al. Examples of suitable packaging virus lines for preparing both ecotropic and amphotropic retroviral systems include Crip, Cre, 2 and

10   Am.

Retroviruses have been used to introduce a variety of genes into many different cell types, including neural cells, epithelial cells, endothelial cells, lymphocytes, myoblasts, hepatocytes, bone marrow cells, in vitro and/or in vivo (see for example Eglitis et al., (1985) Science 230:1395-1398; Danos and Mulligan, (1988) PNAS USA 85:6460-6464; Wilson et al., (1988) PNAS USA 85:3014-

15   3018; Armentano et al., (1990) PNAS USA 87:6141-6145; Huber et al., (1991) PNAS USA 88:8039-8043; Ferry et al., (1991) PNAS USA 88:8377-8381; Chowdhury et al., (1991) Science 254:1802-1805; van Beusechem et al., (1992) PNAS USA 89:7640-7644; Kay et al., (1992) Human Gene Therapy 3:641-647; Dai et al., (1992) PNAS USA 89:10892-10895; Hwu et al., (1993) J. Immunol. 150:4104-4115; U.S. Patent No. 4,868,116; U.S. Patent No. 4,980,286; PCT Application WO

20   89/07136; PCT Application WO 89/02468; PCT Application WO 89/05345; and PCT Application WO 92/07573).

Furthermore, it has been shown that it is possible to limit the infection spectrum of retroviruses and consequently of retroviral-based vectors, by modifying the viral packaging proteins on the surface of the viral particle (see, for example PCT publications WO93/25234, WO94/06920, and

25   WO94/11524). For instance, strategies for the modification of the infection spectrum of retroviral vectors include: coupling antibodies specific for cell surface antigens to the viral env protein (Roux et al., (1989) PNAS USA 86:9079-9083; Julan et al., (1992) J. Gen Virol 73:3251-3255; and Goud et al., (1983) Virology 163:251-254); or coupling cell surface ligands to the viral env proteins (Neda et al., (1991) J. Biol. Chem. 266:14143-14146). Coupling can be in the form of the chemical cross-linking

30   with a protein or other variety (e.g. lactose to convert the env protein to an asialoglycoprotein), as well as by generating fusion proteins (e.g. single-chain antibody/env fusion proteins). This technique, while useful to limit or otherwise direct the infection to certain tissue types, and can also be used to convert an ecotropic vector in to an amphotropic vector.

35

Other Viral Systems

Other viral vector systems that may have application in gene therapy have been derived from herpes virus, e.g., Herpes Simplex Virus (U.S. Patent No. 5,631,236 by Woo et al., issued May 20,

31

1997), vaccinia virus (Ridgeway (1988) Ridgeway, "Mammalian expression vectors," In: Rodriguez R L, Denhardt D T, ed. Vectors: A survey of molecular cloning vectors and their uses. Stoneham: Butterworth,; Baichwal and Sugden (1986) "Vectors for gene transfer derived from animal DNA viruses: Transient and stable expression of transferred genes," In: Kucherlapati R, ed. Gene transfer.

5   New York: Plenum Press; Coupar et al. (1988) Gene, 68:1-10), and several RNA viruses. Preferred viruses include an alphavirus, a poxvirus, an arena virus, a vaccinia virus, a polio virus, and the like. In particular, herpes virus vectors may provide a unique strategy for persistence of the recombinant gene in cells of the central nervous system and ocular tissue (Pepose et al., (1994) Invest Ophthalmol Vis Sci 35:2662-2666). They offer several attractive features for various mammalian cells (Friedmann

10   (1989) Science, 244:1275-1281 ; Ridgeway, 1988, *supra*; Baichwal and Sugden, 1986, *supra*; Coupar et al., 1988; Horwich et al.(1990) J.Virol., 64:642-650).

With the recent recognition of defective hepatitis B viruses, new insight was gained into the structure-function relationship of different viral sequences. *In vitro* studies showed that the virus could retain the ability for helper-dependent packaging and reverse transcription despite the deletion

15   of up to 80% of its genome (Horwich et al., 1990, *supra*). This suggested that large portions of the genome could be replaced with foreign genetic material. The hepatotropism and persistence (integration) were particularly attractive properties for liver-directed gene transfer. Chang et al. recently introduced the chloramphenicol acetyltransferase (CAT) gene into duck hepatitis B virus genome in the place of the polymerase, surface, and pre-surface coding sequences. It was cotransfected with

20   wild-type virus into an avian hepatoma cell line. Culture media containing high titers of the recombinant virus were used to infect primary duckling hepatocytes. Stable CAT gene expression was detected for at least 24 days after transfection (Chang et al. (1991) Hepatology, 14:124A).

Administration of Viral Vectors

25   .   Generally the viral particles are transferred to a biologically compatible solution or pharmaceutically acceptable delivery vehicle, such as sterile saline, or other aqueous or non-aqueous isotonic sterile injection solutions or suspensions, numerous examples of which are well known in the art, including Ringer's, phosphate buffered saline, or other similar vehicles. Delivery of the recombinant viral vector can be carried out via any of several routes of administration, including

30   intramuscular injection, intravenous administration, subcutaneous injection, intrahepatic administration, catheterization (including cardiac catheterization), intracranial injection, nebulization/inhalation or by instillation via bronchoscopy.

Preferably, the DNA or recombinant virus is administered in sufficient amounts to transfect cells within the recipient's target cells, including without limitation, muscle cells, liver cells, various

35   airway epithelial cells and smooth muscle cells, neurons, cardiac muscle cells, etc. and provide sufficient levels of transgene expression to provide for observable ligand-responsive production of a target protein, preferably at a level providing therapeutic benefit without undue adverse effects.

Optimal dosages of DNA or virus depends on a variety of factors, as discussed previously, and

may thus vary somewhat from patient to patient. Again, therapeutically effective doses of viruses are considered to be in the range of about 20 to about 50 ml of saline solution containing concentrations of from about 1×10^7 to about 1×10^{10} pfu of virus/ml, e.g. from 1×10^8 to 1×10^9 pfu of virus/ml.

5   **Uses**

In one application, cells engineered in accordance with the invention are used to produce a target protein in vitro. In such applications, the cells are cultured or otherwise maintained until production of the target protein is desired. At that time, the appropriate ligand is added to the culture medium, in an amount sufficient to cause the desired level of target protein production. The protein so
10   produced may be recovered from the medium or from the cells, and may be purified from other components of the cells or medium as desired.

Proteins for commercial and investigational purposes are often produced using mammalian cell lines engineered to express the protein. The use of mammalian cells, rather than bacteria, insect or yeast cells, is indicated where the proper function of the protein requires post-translational
15   modifications not generally performed by non-mammalian cells. Examples of proteins produced commercially this way include, among others, erythropoietin, BMP-2, tissue plasminogen activator, Factor VIII:c, Factor IX, and antibodies.

In other applications, cells within an animal host or human subject are engineered in accordance with the invention, or cells so engineered are introduced into the animal or human
20   subject, in either case, to prepare the recipient for ligand-mediated regulation of transcription of the target gene. In the case of non-human animals, this can be done as part of veterinary treatment of the animal or to create an animal model for a variety of research purposes. In the case of human subjects, this can be done as part of a therapeutic or prophylactic treatment program.

This invention is applicable to a variety of treatment approaches. For example, the target, e.g.,
25   therapeutic, gene to be regulated can be an endogenous gene or a heterologous gene.

In some cases the therapeutic protein is a factor necessary for the proliferation and/or differentiation of one or more cell types of interest. For example, it may be desirable to stimulate the production of growth factors and lymphokines in a subject in which at least some of the blood cells have been destroyed, e.g., by radiotherapy or chemotherapy. For example, production of
30   erythropoietin stimulates the production of red blood cells, production of G-CSF stimulates the production of granulocytes, production of GM-CSF stimulates the production of various white blood cells, etc. Similarly in diseases or conditions in which one or more specific cell types are destroyed by the disease process, e.g., in autoimmune diseases, the specific cells can be replenished by stimulating production of one or more factors stimulating proliferation of these cells. The method of
35   the invention can also be used to increase the number of lymphocytes in a subject having AIDS, such as by stimulating production of lymphokines, e.g., IL-4, which stimulates proliferation of certain T helper (Th) cells.

Conditional aggregation domains can be used to control a variety of cellular processes that

take place in subcellular compartm nts. These applications involve expression of one or more fusion proteins containing a heterologous domain and a conditional aggregation domain (CAD) in locations such as the nucleus, the cytoplasm (including proteins anchored to the inner leaflet of the plasma membrane), and mitochondria.

In the cases of nuclear and mitochondrial localization, short peptide sequences known in the literature can be added to direct localization. An example is the nuclear localization sequence from the SV40 virus T antigen, P-K-K-K-R-K-V. For location to membranes, a sequence that directs myristoylation or palmitoylation can be added, usually to one or other end of the protein. Such lipid targeting motifs are reviewed in Ann. Rev Biochem (1988) 57, 69. An amino acid sequence of interest

10 includes the sequence M-G-S-S-K-S-K-P-K-D-P-S-Q-R added to the N-terminus. Location in the cytoplasm is believed to be a default state in the absence of any other localization signals.

In these applications the most useful CADs are anticipated to be those known to aggregate through an intrinsic mutual affinity (such as FKBP F36M or W59V). In all these applications, the number of CADs included in the fusion proteins can be varied in order to "tune" the aggregation

15 tendency of the fusion proteins, and the potency with which the aggregation is inhibited by ligands. Increasing the number of CADs will increase the lifetime and affinity of the aggregated complexes by virtue of the avidity effect. A useful model system to assess the degree of aggregation of CAD fusion proteins (eg to optimize the number of CADs to include) is to engineer fusions between CADs and GFP, providing a means to directly visualize the degree of aggregation in living cells. CAD-mediated

20 aggregation of chimeric proteins represents an approach to engineering conditional alleles that is distinct from, yet complementary to, oligomerization of chimeric proteins by small molecule dimerizers as described in WO 94/18317. For CAD-mediated aggregation, the default state is aggregated rather than monomeric, providing a convenient way to investigate the effects of disaggregation. Furthermore the ligands required for disaggregation need only have single binding activity and consequently a

25 single binding surface, in contrast to small molecule dimerizers which are required to bind two proteins simultaneously.

Several types of applications can be envisaged as listed below. These are shown diagrammatically in Figure 1 (A-F).

30

1. A transcriptional off-switch: constitutive transcription of a gene that is turned off by addition of ligand

In these cases, two fusions are constructed and expressed: one between a DNA binding domain and one or more conditional aggregation domain(s), and another between a transcriptional

35 activation domain and one or more conditional aggregation domain(s). These proteins are equipped with nuclear localization sequences. As described in example 2, when these proteins are co-expressed in a cell engineered to contain a gene of interest under the control of DNA binding sites for the DNA binding domain, a functional transcriptional factor will be formed by the non-covalent

linkage of the two fusion proteins via aggregation, and the gene will be transcribed. Addition of a ligand that binds the CAD will inactivate transcription in a dose-dependent manner. Thus, this system can act as a "transcriptional off'switch" (see Clackson, Curr Opin Chem Biol 1997 vol 1, 210). Such an embodiment would be useful, for example, in the case of transgenic animals. If one wanted to

5   determine the effect of a given gene in an adult animal, the animal would be allowed to develop in the absence of ligand and transcription would proceed normally. When the animal reached adulthood, it could then be treated with ligand to turn off expression of the target gene.

Examples of fusion proteins useful for such an application are the chimeric DNA binding protein ZFHD1-(F36M-FKBP)x3 together with the chimeric activation domain (F36M-FKBP)x3-p65,

10 and the chimeric DNA binding protein ZFHD1-(W59V-FKBP)x3 together with the chimeric activation domain (W59V-FKBP)x3-p65. Constitutive transcription by these proteins can be inhibited either by conventional FKBP ligands (eg. FK506, rapamycin, and analogs thereof which are modified in their effector domains to reduce or eliminate immunosuppressive activity while maintaining useful FKBP binding affinity) or by small synthetic ligands designed to bind specifically to these FKBP mutants,

15 which are engineered to contain a side chain truncation.

2.   A transcriptional repression-activation 'toggle' switch: actively repressed transcription of a gene until addition of ligand to displace a chimeric repression domain and recruit activation domain

20 In a specific embodiment of the invention, fusion proteins containing conditional aggregation domains are used to regulate transcription. In this embodiment, nucleic acids encoding two different fusion proteins are introduced into cells. The first fusion protein contains a CAD fused to a DNA binding domain, and the second fusion protein contains a CAD fused to a transcription repression domain. These constructs are introduced into the cell along with a target gene construct operably

25 linked to a constitutive expression control element containing a site recognized by the DNA binding domain of the first fusion protein. In the absence of ligand, the CADs will aggregate and cause repression of the target gene. When ligand is added, the repression is removed and transcription can proceed.

30 **3. Conditional steric inactivation of a transcription factor or other protein: a means to inhibit biological activity of a protein until inhibition is relieved by addition of ligand**

For many proteins, self-association into large aggregates may reduce or eliminate functional activity by, for example, masking an active site or by preventing the protein from interacting with

35 upstream or downstream activators or substrates. Self-association and ligand inducible regulation of a protein's activity may therefore be achieved via the fusion of a CAD or multiple CADs to the protein. In the absence of ligand the activity of the fusion protein may be sharply reduced or eliminated. Dissociation of the aggregates by monomeric ligand would then result in the restoration of the activity

of the protein.

In one example, a CAD or multiple CADs are fused to a transcription factor (consisting of at least a DNA binding domain and a transcriptional activation domain). If aggregation of the transcription factor inhibits its activity, the level of transcription of a target gene that contains binding sites for the
5   transcription factor would be low in the absence of ligand. In the presence of ligand, dissociation of the aggregated transcription factor would lead to the activation of transcription of the target gene.

4. A signalling off-switch

In these applications, CADs are fused to heterologous domains that upon mutual association
10   trigger a signalling event leading to cellular processes such as cell proliferation, differentiation or death. Signalling proteins that are activated by oligomerization at the membrane include the T cell receptor CD3 zeta chain, receptors for cytokines and growth factors such as the EPO receptor, c-kit, the TPO receptor c-mpl, the EGF receptor, etc (ref Klemm et al, Ann Rev Immunol 1998 vol 16, 569). Still other proteins can be activated by oligomerization even when expressed as free cytosolic
15   proteins: for example Raf (Luo et al, Nature 1996 vol 383, 181). Fusion of or more CADs to the signalling domains of these proteins provides a route to small molecule control of these processes. An example of such a fusion protein is myr-(F36M-FKBP)x3-EPORi where myr is an N-terminal myristoylation domain and EPORi is a portion of the intracellular domain of the EPO receptor encompassing the region required for signalling cell proliferation. Expression of such a fusion protein
20   in cultured cells will provide a constitutive proliferation signal by virtue of the F36M-mediated aggregation of the EPORi domains and their consequent activation. This activation can be completely abolished by addition of a cell permable small molecule ligand of F36M-FKBP, such as AP21998.

A number of potential uses for such signalling off-switches can be envisaged. For example, cytokine-dependent cell lines such as BAF-3 have great utility in biological research. BAF-3 cells
25   require IL-3 for growth and survival, but this requirement can be circumvented by providing an alternative proliferation signal from another cytokine receptor. Withholding IL-3 provides a powerful way to assess the signalling potential of molecules such as mutant cytokine receptors that have been introduced into BAF-3 cells by transfection. For example see (Schwaller et al., EMBO J. 1998: 5321-5333). For such applications, a cell line expressing myr-(F36M-FKBP)x3-EPORi could be used
30   instead of BAF-3. These cells would proliferate in the absence of any exogenous cytokine, until a cell permeable F36M ligand were added, at which point survival would require another proliferative signal. Use of such a cell line could have several advantages over BAF-3, including ease of maintenance in culture and cost savings because IL-3 is not needed.

Such a CAD system could also be used as a ligand-inactivatable proliferation switch for
35   hematopoietic cells by analogy to the work of Blau and coworkers (Blau et al PNAS 1997 vol 94, 3076; other Blau references?).

These applications provide a new route to conditional alleles of signalling proteins and other proteins that are activated or inactivated by oligomerization (for review see Spencer, Trends Genet.

1996 vol 12, 181).

5. A method t   study the effects of subcellular localization   n prot in activity.

CAD fusions can be used to assess the effects of abolishing a specific subcellular location of a
5   protein. In these applications, two CAD fusions are expressed: one a fusion between one or more
CADs and a protein domain of interest (but lacking any subcellular localization motifs), and the other an
'anchoring' protein comprising one or more CADs targetted to a particular subcellular location by
addition of a targetting motif as described above. When these proteins are coexpressed in cells, they
will interact by virtue of their CADs, and the protein domain of interest will be restricted to the location
10   dictated by the anchoring protein. Upon addition of CAD ligand, the association between the chimeric
proteins will be disrupted, thereby abolishing the localization of the protein and allowing the effects of
this delocalization on the activities of the protein to be assessed.

An example of such an application is the engineering of a conditionally membrane localized
Raf protein. Coexpression of myr-(F36M-FKBP)x3 and (F36M-FKBP)x3-Raf chimeric proteins inside
15   cells will lead to constitutive membrane localization of Raf. Addition of a cell-permeable F36M ligand
(such as AP21998) will abolish this localization so that F36M-Raf is now distributed throughout the
cell.

Instead of using native proteins of interest (Raf in the example above), dominant negative
mutants could instead be used. Localizing a dominant negative mutant to the same place as an
20   endogenous protein (eg the plasma membrane for Raf and many othe signalling proteins such as Sos,
Grb2 etc) will inhibit activity. This inhibition can be abolished by delocalizing the dominant negative
mutant by adding CAD ligand. The ability to inactivate a dominant negative allele at will using small
molecules allows the real-time consequences of reactivation to be assessed.

25

**6. A method to study the importance of protein-protein interactions by providing a
means to disrupt them with monomer.**

CADs have general utility in cell biological research as a means to study protein-protein
interactions dynamically in living cells. Proteins of particular interest are those that naturally
30   oligomerize, either constitutively or transiently, through discrete domains. These domains can be
replaced with one or more CADs so that the resulting proteins interact constitutively though the CAD
moities when expressed in cells. Addition of ligand for the CAD will dissociate the molecules and allow
the consequences to be determined. An example is the replacing the SH3 domains of signalling
proteins such as PI3K, and their target proline-rich binding sites on PI3K effector proteins, with one or
35   more CADs. These proteins will interact constitutively inside cells, as do the endogenous proteins,
but this interaction can be disrupted at will by adding cell-permeable CAD ligand. This allows the
consequences of abolishing the interaction to be determined in living cells. Such experiments would
be difficult do perform on the endogenous protein-protein complex because of the difficulty of

37

generating cell-permeable specific ligand inhibitors of SH3 domains.

In cases in which the target gene is an endogenous gene of the cells to be engineered, the promoter and/or one or more other regions of the gene can be modified to include a target sequence that is specifically recognized by the DNA binding domain of a fusion protein of this invention so that

5   the endogenous target gene is specifically recognized and regulated in a ligand-dependent manner. Such an embodiment can be useful in situations in which no DNA binding protein is known to specifically bind to a regulatory region of the target gene. Thus, in one embodiment, one or more cells are obtained from a subject or other source and genetically engineered in vitro such that a desired control element is inserted, operatively linked to the target gene. The cell can then be introduced

10   into the subject. Alternatively, prior to introduction of the cell to the subject, the cell is further modified to include a nucleic acid encoding a fusion protein comprising a DNA binding domain which is capable of interacting specifically with the expression control element introduced into the target gene. In other examples of the invention, an endogenous gene is modified in vivo by, e.g., homologous recombination, a technique well known in the art, and described, e.g., in Thomas and Capecchi (1987)

15   Cell 51:503; Mansour et al. (1988) Nature 336:348; and Joyner et al. (1989) Nature 338:153.

Screening for New CADs

Methods to identify proteins that interact with one another are well known. A commonly used

20   technique is the two-hybrid system, in which one partner is fused to a DNA binding domain and the other to an transcriptional activation domain. Interaction of the partners reconstitutes thae transcription factor, activating transcription of a reporter gene that can be identified by screening (eg. production of beta-galactosidase or SEAP) and/or that leads to cell survival and therefore provides a means for selecting for interacting partners (eg. his gene transcription in a his- strain of yeast). Two-hybrid assays

25   can be performed in yeast or mammalian cells and methods are well known in the art.

A preferred embodiment is based on the vectors and cells described by Rivera et al. (Nature Med 1996 2, 1028-1032). Two expression vectors are constructed for chimeric transcription factors in which the candidate CRD is fused to the hybrid DNA domain ZFHD1 (in one case) and to an activiation domain of NF-kB p65 subunit, such as amino acids 361-550 (in the other). These vectors are

30   transiently or stably transfected into mammalian cells, for example HT1080 cells, together with a SEAP reporter gene under the control of ZFHD1 binding sites. Aggregation of the candidates CRDs results in reconstitution of an active transcription factor and therefore prodiction of SEAP. Once a self-aggregating protein has been identified in this way, addition of candidate CRD ligand can be used to examine whether the aggregates can be dissociated with ligand, ie. to test whether the domain is a

35   CAD. Reduction in the production of SEAP upon addition of ligand would indicate this activity. Any polypeptide can be chosen for testing in this way for CAD activity, but preferred proteins to try are those that already have known small molecule binding activity. In these cases the known binding ligands provide a starting point for choosing compounds that might disaggregate bound protein.

As before, an importnat additional configuration to explore is the concatenation of candidate CAD domains. Presence of more than one aggregating domain may increase the apparent affinity of the aggregative interaction by virtue of the avidity effect.

Either natural or mutated proteins can be tested for CAD activity. Mutants of natural proteins are likely to provide good sources of CADs as examples are known on the literature of aggregative activity induced by point mutations: for example sickle-cell hemoglobin, or alpha-1 antitrypsin as described earlier. Thus, large sets of mutants of a candidate protein can be cloned into two-hybrid vectors as described above, and tested for aggregative activity that can be reduced by addition of a small molecule. The criteria that dictate choice of positions to mutate will largely be the same as those described above for screening for CRDs directly in a secretion system (2 above); in addition, mutants that aggregate might be provided by converting polar surface residues to less polar amino ones. Single or multiple mutants can be engineered, using methods as described above.

Selection schemes for CADs can also be devised. In these cases, libraries of mutant proteins are cloned into two hybrid vectors and analyzed en masse for CAD activity. These experiments are most easily performed in yeast and methods for two-hybrid selections are well known in the art. For example, expression vectors for mutants of candidate CADs, fused to GAL4 DNA binding domain or activation domain vectors, and transformed into a his-deficient yeast reporter strain harboring a his gene under the control of GAL4 binding sites. Plating the library on his-deficient medium will result in growth only of cells that contain interacting CADs on the two chimeric transcription factors. These positives can then be replica plated on to plates containing increasing amounts of candidate CAD ligands, to identify those CADs whose interactions can be disrupted by small molecules. Such proteins are candidates for use as CRDs.

A complication with the above selection scheme is the desire to have the same mutant fused to both the DNA binding and activation domains, in order to identify proteins that self-aggregate. To achieve this, the expression vactor for the chimeric proteins can be modified to allow a mutant gene to be joined to both transcription domains at the level of splicing. The domains of interest are encoded in separate exons. An outline of a suitable vector is shown in Fig 5.

CAD cand is the candidate CAD: a library of candidates (eg mutant proteins) is inserted here. DBD and AD are the DNA binding and activation domains of a transcription factor. A and D indicate donor and acceptor splice sites. stop indicates a translational stop codon. By equipping the DBD with a suboptimal splice acceptor site, the CAD exon will be spliced to both DBD and AD exons. Thus, in each cell fusion proteins will be expressed in which the AD and DBD are both fused to an identical CAD candidate.

An alternative format for selection of self-aggregating proteins is the lambda repressor fusion system in E.coli (Hu et al. 1990 Science 250:1400-1403; for review see Hu 1995 Structure 3: 431-433). This strategy utilizes the fact that bacteriophage lambda repressor cI binds to DNA as a homodimer and that binding of such homodimers to operator DNA prevents transcription of phage genes involved in the lytic pathway of the phage life cycle. Thus, bacterial cells expressing functional

lambda repressor are immune to lysis by superinfecting lambda bacteriophage. Respressor protein comprises an amino terminal DNA binding domain (amino acids 1-92) joined by a 40 amino acid flexible linker to a C-terminal dimerization domain. The isolated N-terminal domain binds very weakly to DNA sue to inefficient dimer formation. High affinity DNA binding can be restored by fusing the domain to a

5   heterlogous dimerization domain, such as the GCN4 leucine zipper. A selection system is therefore possible in which phage immunity is used as a selection for interacting proteins.

For example, to select CADs from a library of candidates, the candidates are cloned in frame with the repressor N-terminus and the library transformed into E.coli. Genes for proteins that aggregate are isolated from colonies that survive on plates containing high titers of lambda phage.

10   These colonies can then be restreaked on to plates containing both lambda phage and candidate CAD ligand. If the ligand dissociates the aggregates, the E.coli will now no lolnger grow on these plates. Lambda repressor selection has several advantages for identifying CADs, including the fact that the system is suitable for acreening homodimers, and the large library sizes that can be obtained through the use of E.coli.

15   Another way to directly test whether a protein can act as a CAD in living cells is to fuse its coding sequence to green fluorescent protein (GFP) or variants thereof. Cells expressing such a fusion protein can then be examined directly by fluorescent microscopy to examine whether the CAD candidate appears to cause aggregates of the GFP. Candidate CAD ligand can then be added to determine whether the aggregates then dissociate.

20   **Pharmaceutical Compositions & Their Administration to Subjects Containing Engineered Cells**

Administration

The ligand may be administered to a human or non-human subject using pharmaceutically

25   acceptable materials and methods of administration. Various formulations, routes of administration, dose and dosing schedule may be used for the administration of ligand, depending upon factors such as the condition and circumstances of the recipient, the response desired, the biological half-life and bioavailability of the ligand, the biological half-life and specific activity of the therapeutic protein product, the number and location of engineered cells present, etc. The drug may be administered

30   parenterally, or more preferably orally. For use in this invention, the most preferable route of administration are those in which a rapid onset of response occurs; such methods include, for example, sublingual, buccal, skin patch and inhalation. Dosage and frequency of administration will depend upon factors such as described above. The drug may be taken orally as a pill, powder, or dispersion; buccally; sublingually; injected intravascularly, intraperitoneally, subcutaneously; or the

35   like. The drug may be formulated using conventional methods and materials well known in the art for the various routes of administration. The precise dose and particular method of administration will depend upon the above factors and be determined by the attending physician or healthcare provider.

The particular dosage of the drug for any application may be determined in accordance with

40

conventional approaches and procedures for therapeutic dosage monitoring. A dose of the drug within a predetermined range is given and the patient's response is monitored so that the level of therapeutic response and the relationship of protein production over time may be determined. Depending on the expression levels observed during the time period and the therapeutic response,

5   one may adjust the level of subsequent dosing to alter the resultant expression level over time or to otherwise improve the therapeutic response. This process may be iteratively repeated until the dosage is optimized for therapeutic response. Where the drug is to be administered chronically, once a maintenance dosage of the drug has been determined, one may conduct periodic follow-up monitoring to assure that the overall therapeutic response continues to be achieved.

10   .   In the event that the activation by the drug is to be reversed, administration of drug may be suspended so that cells return to their basal state. To effect a more active reversal of therapy, an antagonist of the drug may be administered. An antagonist is a compound which binds to the drug or drug-binding domain to inhibit interaction of the drug with the fusion protein(s) and thus inhibit the downstream biological event. Thus, in the case of an adverse reaction or the desire to terminate the

15   therapeutic effect, an antagonist can be administered in any convenient way, particularly intravascularly or by inhalation/nebulization, if a rapid reversal is desired.

Compositions

Drugs (i.e., the ligands) for use in this invention can exist in free form or, where appropriate, in

20   salt form. The preparation of a wide variety of pharmaceutically acceptable salts is well-known to those of skill in the art. Pharmaceutically acceptable salts of various compounds include the conventional non-toxic salts or the quaternary ammonium salts of such compounds which are formed, for example, from inorganic or organic acids of bases.

The drugs may form hydrates or solvates. It is known to those of skill in the art that charged

25   compounds form hydrated species when lyophilized with water, or form solvated species when concentrated in a solution with an appropriate organic solvent.

The drugs can also be administered as pharmaceutical compositions comprising a therapeutically (or prophylactically) effective amount of the drug, and a pharmaceutically acceptable carrier or excipient. Carriers include e.g. saline, buffered saline, dextrose, water, glycerol, ethanol, and

30   combinations thereof, and are discussed in greater detail below. The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of

35   mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, etc. Formulation may involve mixing, granulating and compressing or dissolving the ingredients as appropriate to the desired preparation. The pharmaceutical carrier employed may be, for example, either a solid or liquid.

Illustrative solid carriers include lactose, terra alba, sucrose, talc, gelatin, agar, pectin, acacia, magnesium stearate, stearic acid and the like. A solid carrier can include one or more substances which may also act as flavoring agents, lubricants, solubilizers, suspending agents, fillers, glidants, compression aids, binders or tablet-disintegrating agents; it can also be an encapsulating material. In powders, the carrier is a finely divided solid which is in admixture with the finely divided active ingredient. In tablets, the active ingredient is mixed with a carrier having the necessary compression properties in suitable proportions and compacted in the shape and size desired. The powders and tablets preferably contain up to 99% of the active ingredient. Suitable solid carriers include, for example, calcium phosphate, magnesium stearate, talc, sugars, lactose, dextrin, starch, gelatin, cellulose, methyl cellulose, sodium carboxymethyl cellulose, polyvinylpyrrolidine, low melting waxes and ion exchange resins.

Illustrative liquid carriers include syrup, peanut oil, olive oil, water, etc. Liquid carriers are used in preparing solutions, suspensions, emulsions, syrups, elixirs and pressurized compositions. The active ingredient can be dissolved or suspended in a pharmaceutically acceptable liquid carrier such as water, an organic solvent, a mixture of both or pharmaceutically acceptable oils or fats. The liquid carrier can contain other suitable pharmaceutical additives such as solubilizers, emulsifiers, buffers, preservatives, sweeteners, flavoring agents, suspending agents, thickening agents, colors, viscosity regulators, stabilizers or osmo-regulators. Suitable examples of liquid carriers for oral and parenteral administration include water (partially containing additives as above, e.g. cellulose derivatives, preferably sodium carboxymethyl cellulose solution), alcohols (including monohydric alcohols and polyhydric alcohols, e.g. glycols) and their derivatives, and oils (e.g. fractionated coconut oil and arachis oil). For parenteral administration, the carrier can also be an oily ester such as ethyl oleate and isopropyl myristate. Sterile liquid carders are useful in sterile liquid form compositions for parenteral administration. The liquid carrier for pressurized compositions can be halogenated hydrocarbon or other pharmaceutically acceptable propellant. Liquid pharmaceutical compositions which are sterile solutions or suspensions can be utilized by, for example, intramuscular, intraperitoneal or subcutaneous injection. Sterile solutions can also be administered intravenously. The drugs can also be administered orally either in liquid or solid composition form.

The carrier or excipient may include time delay material well known to the art, such as glyceryl monostearate or glyceryl distearate along or with a wax, ethylcellulose, hydroxypropylmethylcellulose, methylmethacrylate and the like. When formulated for oral administration, 0.01% Tween 80 in PHOSAL PG-50 (phospholipid concentrate with 1,2-propylene glycol, A. Nattermann & Cie. GmbH) may be used as an oral formulation for a variety of drugs for use in the practice of this invention.

A wide variety of pharmaceutical forms can be employed. If a solid carrier is used, the preparation can be tableted, placed in a hard gelatin capsule in powder or pellet form or in the form of a troche or lozenge. The amount of solid carrier will vary widely but preferably will be from about 25 mg to about 1 g. If a liquid carrier is used, the preparation will be in the form of a syrup, emulsion, soft gelatin capsule, sterile injectable solution or suspension in an ampule or vial or nonaqueous liquid

suspension.

To obtain a stable water soluble dosage form, a pharmaceutically acceptable salt of the drug may be dissolved in an aqueous solution of an organic or inorganic acid, such as a 0.3M solution of succinic acid or citric acid. Alternatively, acidic derivatives can be dissolved in suitable basic solutions. If

5   a soluble salt form is not available, the compound is dissolved in a suitable cosolvent or combinations thereof. Examples of such suitable dissolved in a suitable cosolvent or combinations thereof. Examples of such suitable cosolvents include, but are not limited to, alcohol, propylene glycol, polyethylene glycol 300, polysorbate 80, glycerin, polyoxyethylated fatty acids, fatty alcohols or glycerin hydroxy fatty acids esters and the like in concentrations ranging from 0-60% of the total

10   volume.

Various delivery systems are known and can be used to administer the drugs, or the various formulations thereof, including tablets, capsules, injectable solutions, encapsulation in liposomes, microparticles, microcapsules, etc. Preferred routes of administration to a patient are oral, sublingual, transdermal (patch), intranasal, pulmonary or bucal. Methods of introduction also could include but are

15   not limited to dermal, intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, epidural, ocular and (as is usually preferred) oral routes. The drug may be administered by any convenient or otherwise appropriate route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (e.g., oral mucosa, rectal and intestinal mucosa, etc.) and may be administered together with other biologically active agents. Administration can be systemic or local.

20   For ex vivo applications, the drug will be delivered as a liquid solution to the cellular composition.

In a specific embodiment, the composition is formulated in accordance with routine procedures as a pharmaceutical composition adapted for intravenous administration to human beings. Typically, compositions for intravenous administration are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic to ease

25   pain at the side of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the composition is administered by

30   injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

In addition, in certain instances, it is expected that the compound may be disposed within devices placed upon, in, or under the skin. Such devices include patches, implants, and injections which release the compound into the skin, by either passive or active release mechanisms.

35   Materials and methods for producing the various formulations are well known in the art and may be adapted for practicing the subject invention. See e.g. US Patent Nos. 5,182,293 and 4,837,311 (tablets, capsules and other oral formulations as well as intravenous formulations) and European Patent Application Publication Nos. 0 649 659 (published April 26, 1995; rapamycin

formulation for IV administration) and 0 648 494 (published April 19, 1995; rapamycin formulation for oral administration).

The effective dose of the drug will typically be in the range of about 0.01 to about 50 mg/kgs, preferably about 0.1 to about 10 mg/kg of mammalian body weight, administered in single or multiple

5   doses. Generally, the compound may be administered to patients in need of such treatment in a daily dose range of about 1 to about 2000 mg per patient. In embodiments in which the compound is rapamycin or an analog thereof with some residual immunosuppressive effects, it is preferred that the dose administered be below that associated with undue immunosuppressive effects.

The amount of a given drug which will be effective in the treatment or prevention of a particular

10   disorder or condition will depend in part on the severity of the disorder or condition, and can be determined by standard clinical techniques. In addition, in vitro or in vivo assays may optionally be employed to help identify optimal dosage ranges. Effective doses may be extrapolated from dose-response curves derived from in vitro or animal model test systems. The precise dosage level should be determined by the attending physician or other health care provider and will depend upon

15   well known factors, including route of administration, and the age, body weight, sex and general health of the individual; the nature, severity and clinical stage of the disease; the use (or not) of concomitant therapies; and the nature and extent of genetic engineering of cells in the patient.

The drugs can also be provided in a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions. Optionally

20   associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceutical or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

                   ■   ■   ■   ■   ■

25

The full contents of all references cited in this document, including references from the scientific literature, issued patents and published patent applications, are hereby expressly incorporated by reference. In addition, the full contents of Rothman, Clackson and Rivera, "Methods and Materials Involving Conditional Retention Domains" filed October 19, 1998 (Attorney docket No.

30   ARIAD 383 US) are expressly incorporated herein by reference.

The following examples contain important additional information, exemplification and guidance which can be adapted to the practice of this invention in its various embodiments and the equivalents thereof. The examples are offered by way of illustration only and should not be construed as limiting in any way. As noted throughout this document, the invention is broadly applicable and

35   permits a wide range of design choices by the practitioner.

The practice of this invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, immunology, virology, pharmacology, chemistry, and pharmaceutical formulation

and administration which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No:

5   4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor

10   Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.), Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

15   **Examples**

Example 1:Generation of domains and vectors used for expression of F(36M) fusion proteins.

20   **A. Expression vectors:**

Vectors for driving expression of fusion proteins were derived from the mammalian expression vector pCGNN (Attar and Gilman, MCB 12:2432-2443, 1992). Inserts cloned as XbaI-BamHI fragments into pCGNN are transcribed under the control of the human CMV promoter and enhancer sequences

25   (nucleotides -522 to +72 relative to the cap site), and are expressed with an N-terminal nuclear localization sequence (NLS; from SV40 T antigen) and epitope tag (a 16 amino acid portion of the *H. influenzae* hemaglutinin gene).

pCGNN was modified by site directed mutagenesis with oligonucleotides VR65, VR119, and VR120

30   to create pC4EN. The resulting plasmid has unique restriction sites upstream of the CMV enhancer/promoter region (MluI) and between the promoter and protein coding region (EcoRI).

VR65: TCCCGCACCTCTTCGGCCAGCGaaTTccAGAAGCGCGTAT
VR119: GACTCACTATAGGaCGcgTTCGAGCTCGCCCC

35   VR120: CATCATTTTGGCAAAGgATTCACTCCTCAGG

Individual components of fusion proteins were generally produced as fragments containing an XbaI site immediately upstream of the first codon and a SpeI site, an in-frame stop codon, and a BamHI site

immediately downstream of the last codon. Chimeric proteins comprising multiple components were assembled by stepwise insertion of XbaI-BamHI fragments into SpeI-BamHI-opened vectors or by insertion of XbaI-SpeI fragments into XbaI or SpeI-opened vectors.

5

B. F(36M) domain

F(36M), in which the phenylalanine at amino acid 36 was changed to methionine, was created by

10   mutagenizing a single FKBP domain, cloned into pCGNN with upstream XbaI and downstream SpeI and BamHI sites (Rivera et al., Nat. Med 2:1028-1032, 1996) with oligo VR1 to create pCGNN-F(36M). Two, 3, 4 and 6 tandem copies of F(36M) were created by the stepwise insertion of XbaI-BamHI fragments into SpeI-BamHI-opened vectors.

15   VR1: GATGGAAAGAAAatgGATTCCTCCCGG

C. F(36M) fusion proteins

(a) EGFP fusions

20   EGFP coding sequence was amplified from pEGFP-1 (Clontech) with oligos VR2 and VR3. The resulting fragment, with upstream XbaI and downstream SpeI sites was inserted into pCGN, a derivative of pCGNN that lacks the SV40 nuclear localization sequence, to create pCGN-EGFP.

VR2: tctagaGTGAGCAAGGGCGAGGAG

25   VR3: ggatccttaTTAACTAGTCTTGTACAGCTCGTCCATG

F(36M)-EGFP fusions were created by inserting XbaI-SpeI fragments containing 3, 4 or 6 copies of F(36M) into the XbaI site of pCGN-EGFP to create pCGN-F(36M)3-EGFP, pCGN-F(36M)4-EGFP, and pCGN-F(36M)6-EGFP*.

30

(b) hGH fusions

An hGH cDNA (506-81) was obtained by RT-PCR amplification of RNA expressed from a cell line containing a genomic hGH gene (Rivera et al., Nat. Med 2:1028-1032, 1996) using oligos VR109 and

35   VR110 to amplify the region from 40 bp upstream of the ATG to 60 bp after the stop codon. The resulting HindIII to EcoRI fragment was cloned into Z12I-PL-2, a derivative of ZHWTx12-IL2-SEAP (Rivera et al., Nat. Med 2:1028-1032, 1996) in which the SEAP gene and SV40 early intron and polyadenylation signal were replaced by a polylinker and the SV40 late polyadenylation signal.

VR109: aagcttACCACTCAGGGTCCTGTGG

VR110: gaattcGTGGCAACTTCCA

5   To construct hGH fusion proteins, Z12I-hGH-2 was mutagenized with oligos VR185, VR186, and
     VR187 to create i) an EcoRI site 32 bp upstream of the ATG, ii) an XbaI site immediately after the last
     amino acid of the signal sequence and iii) a Spe I site immediately after the last amino acid of hGH.

     VR185: cacaggaccctGAATTCtaagcttgtggc

10   VR186: ATAAGGGAATGGTtctagaGGCACTGCCCT
     VR187: atgccacccgggactagtGAAGCCACAGCTG

     Cloning the resulting EcoRI-SpeI fragment into pC4EN produced pC4S1-hGH which expresses hGH
     from the CMV enhancer. The XbaI-BamHI fragment of pC4S1-hGH was then replaced by XbaI-SpeI

15   fragments containing 2, 3, 4, or 6 copies of F(36M) and a SpeI-BamHI fragment encoding the furin
     cleavage site-hGH fusion to generate pC4S1-F(36M)-FCS-hGH fusions.

     A SpeI-BamHI fragment encoding an FCS-hGH fusion protein was generated by amplification of the
     hGH cDNA with oligos VR4 and VR5.

20

     VR4:actagtGCTAGAAACCGTCAGAAGAGATTCCCAACCATTCCCTTAAGC
     VR5: ggatcccgggCTAGAAGCCACAGCTGCCCTC

     An XbaI-BamHI fragment containing the neo resistance gene downstream of the

25   encephalomyocarditis virus internal ribosome entry sequence (IRES/Neo; Amara et al PNAS
     94:10618-23, 1997) was inserted into appropriate SpeI-BamHI-opened vectors to generate
     pC4S1-F(36M)-FCS-hGH/neo and pC4S1- hGH/neo vectors.

30   **(c) Insulin fusions**
     A human insulin cDNA was obtained by RT-PCR amplification of human pancreas polyA+ RNA
     (Clontech) using oligos VR220 and VR221 to amplify the region from 9 bp upstream of ATG (EcoRI) to
     13 bp after stop codon (BamHI). The resulting EcoRI-BamHI fragment was cloned into pC4EN to
     generate pC4-hIn.

35

     VR220: cGAATTCttctgccATGGCCCTGTGGATGCGC
     VR221: cGGATCCgcaggctgcgtCTAGTTGCAGTAG

A SpeI-BamHI fragment encoding an furin cleavag sequence-insulin fusion protein was generated by RT-PCR amplification with oligos VR222 and VR221.

VR222: cACTAGTGCTAGAAACCGTCAGAAGAGATTTGTGAACCAACACCTGTGCGGC

5   VR221: cGGATCCgcaggctgcgtCTAGTTGCAGTAG

The wild type insulin gene and FCS-insulin fusion were mutagenized to i) alter amino acid B10 to Asp, ii) create a FCS at the B-C junction, and iii) create a FCS at the C-A junction, using oligos VR223, VR224, VR225, respectively.

10

VR223: CCTGTGCGGCTCAgACCTGGTGGAAGC
VR224: CTTCTACACACCCAgGACCaagCGGGAGGCAGAGG
VR225: CCCTGGAGGGGTCCCgGCAGAAGCGTGGC

15   Mutation of pC4-hIn produced pC4-hIn-m3. The mutated FCS-insulin fusions were used to replace the FCS-hGH portion of the pC4S1-F(36M)-FCS-hGH fusions to create pC4S1-F(36M)-FCS-hIn-m3 fusions.

(d) LNGFR fusions

20   EcoRI-SpeI fragments containing amino acids 1-274 of the human low affinity nerve growth factor receptor (LNGFR; Clackson et al., PNAS 95:10437-42, 1998) and SpeI-BamHI fragments containing 3, 4, or 6 copies of F(36M) were cloned into pC4EN to generate pC4LNGFR-F(36M) fusions.

25   **(c) Transcription factor fusions**

pCGNN-ZFHD1-F(36M) and pCGNN-F(36M)-p65 fusion proteins were generated as described for wild type FKBP fusions (Amara et al PNAS 94:10618-23, 1997).

An XbaI-SpeI fragment containing 6 copies of F(36M) was inserted into the XbaI or SpeI site of
30   pCGNN-ZFHD1-p65 (reference earlier patent?) to generate pCGNN- F(36M)6-ZFHD1-p65 and pCGNN-ZFHD1-p65-F(36M)6.

Example 2: Identification of F36M FKBP as a conditional aggregation domain.

A candidate CAD was identified during the course of studies aimed at identifying mutant
35   FKBP12 (FKBP) domains that could bind modified FKBP ligands.

The goal was to modify an FKBP ligand such as AP1510 (a synthetic version of FK506) so that it could not interact with wild type, endogenous FKBP, and then modify FKBP so that it could interact with the modified drug (bumps and holes; see Clackson et al PNAS 95:10437-10442 1998). Modified

drug/FKBP combinations were screened by fusing multiple FKBP domains to both a DNA binding domain and a transcription activation domain and measuring expression of a SEAP target reporter gene upon transient transfection into cells. Since FKBP domains have no natural affinity for one another, transcription would only be expected to occur in the presence of a dimeric FKBP ligand (see

5   Fig 2A; Amara et al., PNAS 94:10618-10623 1997; see also figure 2A, infra). However, with one particular FKBP mutant, in which Phe at amino acid 36 was replaced by a Met [F(36M)], the target gene was transcribed constitutively. In the experiment shown in Fig. 2A, Wild-type, F36V or F36M FKBP were fused in one or three copies to the DNA binding domain ZFHD1 or the activation domain p65(361-550) as described (Pollock and Rivera, Meth. Enz. 306:263-281, 1999). Constructs were

10   cotransfected into mammalian cells and tested for their ability to interact in a two-hybrid assay. As expected, wild-type fusion proteins did not interact except in the presence of the dimerizer AP1510 (Amara et al., Proc. Natl. Acad. Sci. USA 94:10618-10623, 1997). The only combinations that interacted were those of F36M fusion proteins, with the strongest interaction being between two (F36M x 3) fusion proteins. Transcription resulting from this interaction was as robust as a control

15   construct comprising covalently fused ZFHD1 and p65, attesting to the robustness of the interaction. Moreover, transcription induced by these fusion proteins could be inhibited by monomeric FKBP ligands such as FK506 or "bumped" synthetic ligands such as AP21998 and AP22542 (Fig 2B). These results suggested that the F(36M) mutant version of FKBP self-associates and that the interaction can be blocked by a monomeric ligand, as illustrated in Figure 1A .

20   To test further the possibility that F(36M) may function as a CAD, varying numbers of F(36M) domains were fused to EGFP and expression constructs transiently transfected into cells. In contrast to EGFP protein alone, which was distributed throughout the cytoplasm, F(36M)-EGFP fusions had a much more punctate expression pattern, with the EGFP protein visible as large green spots concentrated at multiple points throughout the cytoplasm. This suggests that F(36M) domains can

25   induce aggregation of a fused heterologous protein. The degree of aggregation correlated with the number of F(36M) domains fused to EGFP -- when 3 copies were fused, significant amounts of "free" green protein were still visible throughout the cell whereas when 4 or, especially, 6 copies were fused, almost all of the EGFP protein appeared to be aggregated. Consistent with the notion that the punctate expression pattern reflects the formation of aggregates of the fusion protein, it was found

30   that in the presence of FK506 the fusion protein becomes evenly distributed throughout the cell. Furthermore, the ligand-induced disaggregation is rapid and reversible: FK506 disrupts the aggregates in just 5-10 minutes and, if FK506 is washed away, aggregates reform quickly.

Example 3: Identification and synthesis of ligands for the conditional
35   **aggregation domain F36M-FKBP.**

     AP21998 and AP22542 are ligands of FKBP that have particular utility for CAD applications, because they bind with high affinity to F36M-FKBP but poorly to the wild-type protein, and are thus anticipated to lead to minimal interactions with the endogenous proteins during in vivo applications.

The design and assay of such "bumped" ligands that target a hole created by truncating FKBP residue Phe36 have been described (Clackson et al., Proc. Natl. Acad. Sci. USA 95:10437-10442, 1998).

AP 21998 was prepared via DCC/DMAP-mediated coupling of the previously described acid AP 1867 (compound 5S in Clackson et al., Proc. Natl. Acad. Sci. USA 95:10437-10442, 1998) with

5   commercially available N,N-dimethyl-1,3-propanediamine (Scheme 1). AP 22542 was also synthesized by a DCC/DMAP-mediated coupling of acid AP 17362 with alcohol 3 (Scheme 2). Carbinol 3 itself was prepared via a three step sequence as outlined in Scheme 2. The Claisen-Schmidt condensation of 3,4-dimethoxybenzaldehyde and 3-acetylpyridine provided unsaturated ketone 1 as a crystalline solid in 68% yield. Transfer hydrogenation of 1 utilizing ammonium formate as

10   a hydrogen source provided the propanone adduct 2 as a crystalline solid in 50% isolated yield. Finally, the enantioselective reduction of the aryl ketone moiety of 2 to the desired R-configured carbinol 3 was achieved in 86% by reduction of 2 with (+)-b-chlorodiisopinocamphenylborane (DIP-Chloride™) (Chandrasekharan et al. J. Org. Chem. 50:5446, 1985). The synthesis of the acid component, AP 17362, was prepared as described in Scheme 3. The commercially available 3,4,5-

15   trimethoxyphenylacetic acid was converted to the racemic 2-arylbutane derivative 4 in 83% yield by alkylation with iodoethane of the NaHMDS-generated dianion of 3,4,5-trimethoxyphenylacetic acid in THF at 0 oC. Resolution of the acid by repetitive crystallization of the (-)-cinchonidine salt afforded optically enhanced 4S in 24% yield (48% theoretical) and of 91% ee. This resolved acid was then coupled with methyl-L-pipecolate hydrochloride by use of 2-chloro-1-methylpyridinium iodide

20   (Mukaiyama's Reagent). The resulting coupled product was not isolated, but subjected to hydrolysis to afford the desired crystalline acid, AP 17362, in 42% overall yield and >99% de. X-ray structural analysis confirmed the absolute stereochemistry of the resolved 2-arylbutane center as the S configuration.

DCC, DMAP

diamine, CH₂Cl₂, 0 °C
39%

AP 1867   AP 21998

25
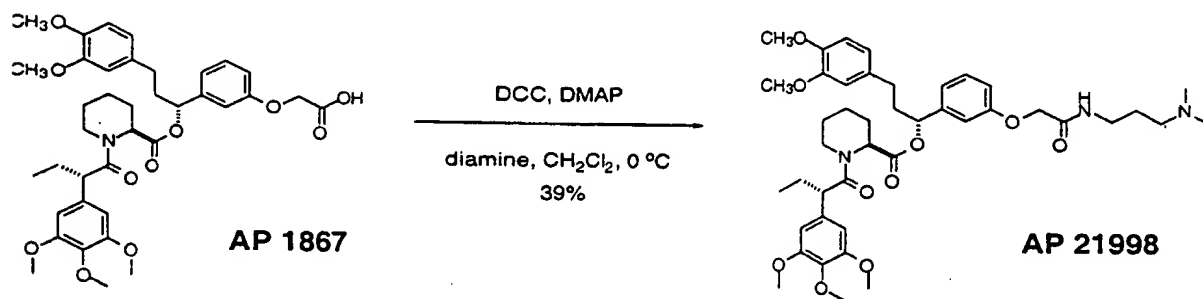
SCHEME 1

AP 21998: A solution of AP 1867 (5.0 g, 7.21 mmol) in CH2Cl2 (5.0 mL) at 0 °C was treated with DCC (178 mg, 0.79 mmol) followed 30 min later by N,N-dimethyl-1,3-propanediamine (880 mg, 8.65 mmol) and DMAP (5 mg). The reaction mixture was allowed to warm to room temperature and stir for 5 h, after which time the reaction mixture was diluted with EtOAc (50 mL), filtered, and the filtrate

5   extracted with a 5% aqueous citric acid solution (3 x 20 ml). The acid extract was then made basic by the addition of solid NaHCO3 and extracted with EtOAc (3 x 50 mL). The organic extract was dried over Na2SO4, filtered, and evaporated to afford a crude material which was flash chromatographed on silica gel (5% then 15% MeOH/ CH2Cl2) to afford product (2.2 g, 39%) as a colorless foam: IR (neat) 2940, 1735, 1650, 1510, 1460, 1240, 1130 cm^{-1}; ^{1}H NMR (CDCl3, 300 MHz) 7.78 (br t, J= 5.1 Hz, 1

10 H), 7.19 (t, J= 8.6 Hz, 1 H), 6.92-6.65 (m, 6 H), 6.42 (s, 2 H), 5.63 (dd, J= 8.0, 5.5 Hz, 1 H), 5.45 (d, J= 4.1 Hz, 1 H), 4.49 (s, 2 H), 3.86-3.70 (m, 16 H), 3.60 (t, J= 7.0 Hz, 1 H), 3.47-3.41 (m, 2 H), 2.82 (td, J= 13.2, 2.4 Hz, 1 H), 2.62-2.29 (m, 12 H), 2.16-1.23 (m, 10 H), 0.90 (t, J= 7.3 Hz, 3 H); ^{13}C NMR (CDCl3, 75 MHz) 172.7, 170.6, 168.5, 157.5, 153.2, 148.9, 147.4, 142.3, 136.7, 135.3, 133.4, 129.8, 120.2, 119.6, 113.9, 112.8, 111.8, 111.4, 105.1, 75.7, 67.3, 60.8, 56.3, 56.0, 52.1, 50.7, 44.3,

15 43.5, 38.3, 37.4, 31.3, 28.3, 26.8, 25.5, 25.4, 20.9, 12.5; LRMS (ES+): (M+H)$^+$ 778; HRMS (FAB): (M+H)$^+$ calcd: 778.4278, meas: 778.4299.
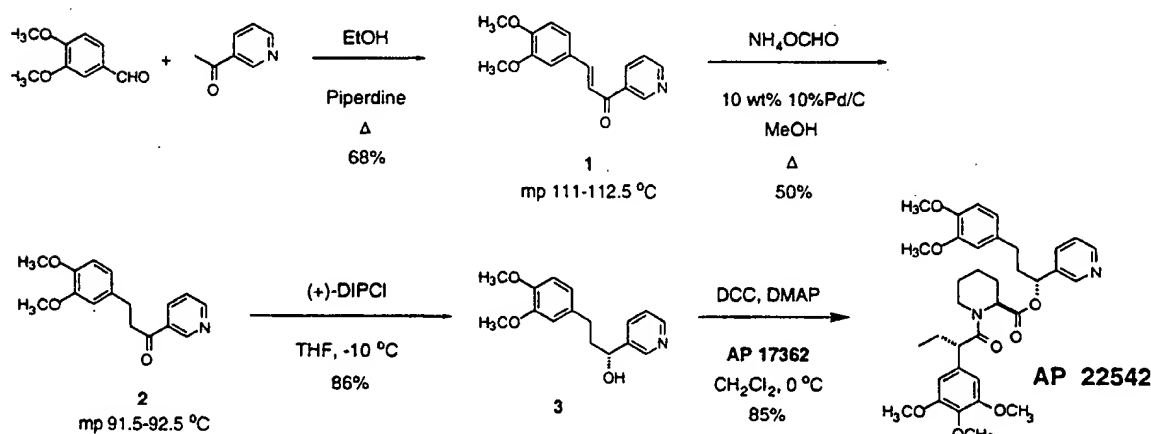

20 SCHEME 2

(E)-3-(3,4-Dimethoxyphenyl)-1-pyridin-3-yl-propenone (1): A solution of 3,4-dimethoxybenzaldehyde (53.7 g, 323 mmol) and 3-acetylpyridine (39.1 g, 323 mmol) in EtOH (400

mL) was treated with piperdine (4.75 mL, 48 mmol) and heated at reflux for 4 days. The reaction was then evaporated to a slurry and treated with water (400 mL). The resulting solids were filtered, air dried, and recrystallized from EtOAc/hexane to afford product (59.2 g, 68%) as a yellow colored solid: mp 111-112.5 °C; TLC (EtOAc) Rf = 0.30; 1H NMR (CDCl3, 300 MHz) 9.23 (d, J= 1.8 Hz, 1 H), 8.79 (dd,

5   J= 4.8, 1.7 Hz, 1 H), 8.28 (dt, J= 7.9, 1.9 Hz, 1 H), 7.79 (d, J= 15.6 Hz, 1 H), 7.46-7.42 (m, 1 H), 7.35 (d, J= 15.6 Hz, 1 H), 7.24 (dd, J= 8.3, 1.9 Hz, 1 H), 7.68 (d, J= 1.9 Hz, 1 H), 6.91 (d, J= 8.3 Hz, 1 H), 3.95 (s, 3 H), 3.93 (s, 3 H); 13C NMR (CDCl3, 75 MHz) 189.0, 152.9, 151.9, 149.7, 149.4, 146.1, 135.8, 133.8, 127.5, 123.6, 119.4, 111.2, 110.2, 56.0 ; LRMS (ES+) (M+H)+ 270; Anal. Calcd for C16H15NO3: C, 71.36; H, 5.61; N, 5.20. Found: C, 71.13; H, 5.70; N, 4.95.

10   3-(3,4-Dimethoxyphenyl)-1-pyridin-3-yl-propan-1-one (2): A solution of olefin 1 (20.0 g, 74.2 mmol), wet 10% Pd/C (2.0 g), and ammonium formate (14.0 g, 222 mmol) in MeOH (400 mL) was heated at reflux for 30 min and filtered, while hot, through a pad of Celite. The filtrate was allowed to slowly cool and the resulting solids were filtered and air dried to afford product (10.0 g, 50%) as a colorless solid: mp 91.5-92.5 °C; TLC (EtOAc) Rf = 0.55; 1H NMR (CDCl3, 300 MHz) 9.16 (d, J= 2.0 Hz, 1 H), 8.76

15   (dd, J= 4.8, 1.7 Hz, 1 H), 8.21 (dt, J= 8.0, 1.9 Hz, 1 H), 7.40 (dd, J= 7.9, 4.8 Hz, 1 H), 6.83-6.77 (m, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.30 (d, J= 7.3 Hz, 2 H), 3.03 (d, J= 7.7 Hz, 2 H); 13C NMR (CDCl3, 75 MHz) 198.2, 153.5, 149.6, 149.0, 147.6, 135.3, 133.4, 132.1, 123.6, 120.2, 111.9, 111.5, 56.0 (2), 40.9, 29.5; Anal. Calcd for C16H17NO3: C, 70.83; H, 6.32; N, 5.16. Found: C, 70.63; H, 6.42; N, 5.05.

20

(R)-3-(3,4-Dimethoxyphenyl)-1-pyridin-3-yl-propan-1-ol (3): A solution of (+)-DIP-Chloride™ (7.09 g, 22.1 mmol) in THF (10 mL) at -25 °C was treated with ketone 2 (2.0 g, 7.37 mmol). The resulting mixture was allowed to stand in at -20 °C for 2 h then placed in a -10 °C freezer for 48 h, after which time the mixture was concentrated and treated with diethyl ether (50 mL) followed by diethanolamine (4.24

25   mL, 44.2 mmol). The viscous mixture was allowed to stir at room temperature for 6 h after which time it was filtered through a pad of Celite with the aid of diethyl ether. The filtrate was concentrated and the crude material flash chromatographed (EtOAc then 10% MeOH/EtOAc) to afford product. The product was redissolved in diethyl ether (50 mL) and again treated once again with diethanolamine (2.12 mL, 22.1 mmol) as described above to afford product (1.74 g, 86%) as a clear colorless oil (96% ee by

30   Chiralpak AD HPLC, 15% EtOH/hexane, retention time 6.1 min for the S-enantiomer and 19.4 min for the desired R-enantiomer): TLC (EtOAc) Rf = 0.25; IR (neat) 3210, 2935, 1590, 1515, 1465, 1420, 1260, 1155, 1070, 1030, 1030 cm-1; 1H NMR (CDCl3, 300 MHz) 8.50 (d, J= 1.7 Hz, 1 H), 8.44 (dd, J= 4.7, 1.5 Hz, 1 H), 7.71 (dt, J= 7.8, 1.7 Hz, 1 H), 7.28-7.24 (m, 1 H), 6.80-6.70 (m, 1 H), 4.72 (dd, J= 7.9, 5.2 Hz, 1 H), 3.85 (s, 6 H), 3.21 (br s, 1 H), 2.77-2.9 (m, 2 H), 2.18-1.96 (m, 2 H); 13C NMR

35   (CDCl3, 75 MHz) 149.0, 148.6, 147.7, 147.4, 140.3, 134.0, 133.8, 123.6, 120.2, 111.8, 111.4, 71.3, 56.0, 55.8, 40.7, 31.5; LRMS (ES+) (M+H)+ 274; HRMS (ES+): (M+H)+ calcd: 274.1462, meas: 274.1443.

1-[2(S)-(3,4,5-trimethoxyphenyl)-butyryl]-piperdine-2(S)-carboxylic acid, 3-(3,4-Dimethoxyphenyl)-1-pyridin-3-yl-propan-1(R)-ol ester (AP22542): A solution of alcohol 3 (600 mg, 2.20 mmol), acid AP17362 (882 mg, 2.42 mmol), and DMAP (2.41 mg, 1.98 mmol) in CH2Cl2 (2.5 mL) at -10 °C, was treated with DCC (498 mg, 2.42 mmol). The mixture was allowed to warm to ~5 °C over a 1 h period

5

10
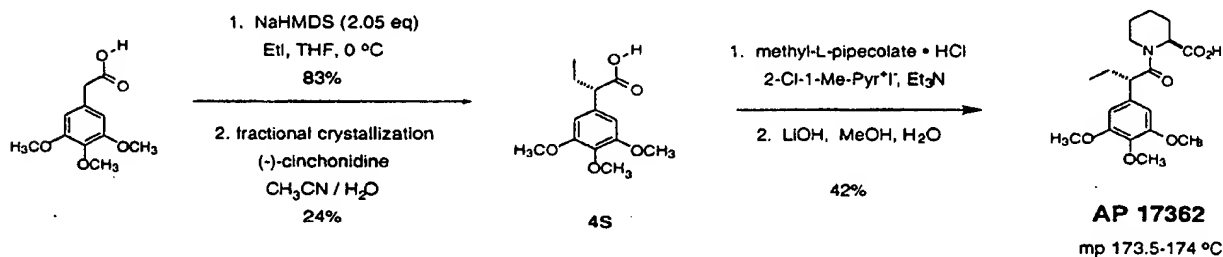
and then placed in a 5 °C refrigerator for an additional 16 h. The reaction mixture was then diluted with EtOAc (3 mL), filtered, evaporated, and the crude material flash chromatographed (75% then 100% EtOAc/hexane) to afford product (1.15 g, 85%) as a colorless foam: TLC (EtOAc) Rf = 0.40; IR (neat) 2940, 1740, 1645, 1590, 1515, 1455, 1420, 1240, 1130, 1030 cm-1; 1H NMR (CDCl3, 300 MHz) 8.50 (dd, J= 4.6, 1.5 Hz, 1 H), 8.42 (d, J= 1.7 Hz, 1 H), 7.27 (d, J= 8.6 Hz, 1 H), 7.19 (dd, J= 7.7, 4.7 Hz, 1 H), 6.78 (d, J=7.7 Hz, 1 H), 6.66-6.64 (m, 2 H), 6.46 (s, 2 H), 5.69 (dd, J= 7.7, 6.0 Hz, 1 H), 5.47 (d, J= 4.3 Hz, 1 H), 3.86-3.73 (m, 16 H), 3.59 (t, J= 7.1 Hz, 1 H), 2.72 (td, J= 13.2, 2.6 Hz, 1 H), 2.60-2.38 (m, 2 H), 2.30 (d, J= 12.4 Hz, 1 H), 2.16-2.02 (m, 2 H), 1.99-1.90 (m, 1 H), 1.79-1.57 (m, 4 H), 1.46-1.37 (m, 1 H), 1.32-1.19 (m, 1 H), 0.90 (t, J= 7.3 Hz, 3 H); 13C NMR (CDCl3, 75 MHz) 172.6, 170.5, 153.3, 149.5, 149.0, 148.3, 147.5, 136.9, 135.6, 135.3, 133.8, 1323.0, 123.6, 120.2, 111.7, 111.5, 105.1, 73.6, 60.9, 56.1, 56.0, 52.0, 50.7, 43.5, 37.9, 31.1, 28.3, 26.7, 25.3, 20.9, 12.5; LRMS (ES+) (M+H)+ 621; HRMS (FAB): (M+H)+ calcd: 621.3176, meas: 621.3178.

25

SCHEME 3

(R/S)-2-(3,4,5-Trimethoxyphenyl)butyric acid: A solution of of 3,4,5-trimethoxyphenylacetic acid (40.0 g, 176.8 mmol) in THF (125 mL) at 0 °C was treated dropwise with a 2N THF solution of sodium bis(trimethylsilyl)amide (181 mL, 362 mmol, Lancaster) over a 1 h period keeping the internal reaction temperature below 8 °C. After 15 min, iodoethane (14.9 mL, 185.7 mmol) was added slowly over a 30 min period keeping the internal reaction temperature below 6-8 °C and the solution allowed to warm to room temperature. After 2 h, the mixture was poured onto EtOAc (700 mL) and acidified by slow addition of a 2.0 N HCl solution (325 mL). The organic component was further washed with a saturated sodium bisulfite solution (50 mL) followed by brine (2 x 50 mL), then dried over anhydrous Na2SO4, and concentrated to a waxy residue (43.8 g). The crude product was recystallized from hot EtOAc/hexane (30 mL/30 mL) to afford product (37.1 g, 83%): mp 103-104 °C; TLC

(AcOH/EtOAc/hexane, 2:49:49) Rf = 0.50.

(S)-2-(3,4,5-Trimethoxyphenyl)butyric acid (4S): A solution of 4 (3.09 g, 12.15 mmol) in CH3CN (130 mL) was treated with (-)-cinchonidine (3.58 g, 12.15 mmol) and the mixture heated to reflux. The homogeneous solution was allowed to slowly cool to room temperature with concomitant formation of salts. After a period of 1 h at room temperature, the solution was cooled to 0˚C for 30 minutes and the solution then filtered to afford 4.05 g of a chalky colorless solid. This recrystalliztion procedure was then carried out an addition four times utilizing ~20 mL CH3CN/g of salt. The diastereomeric salt isolated from the fifth crystallization (1.64 g) was suspended in EtOAc (100 mL) and treated with a 10% aqueous HCl solution (10 mL). The organic phase was then washed with water (2 x 15 mL) followed by brine 10 mL), dried over anhydrous MgSO4, and concentrated to afford product (0.75 g, 24%) as a colorless solid (91% ee by Chiralcel OD HPLC, 1:5:94 formic acid/i-PrOH/hexane, retention time 19.6 min for the R-enantiomer, and 22.1 min for the desired S-enantiomer): mp 84-85 ˚C (99.1% ee material); [a]22D +54.8 (c = 1.07, MeOH, 30 min, 99.1% ee material); UV (MeOH) lmax 270 (e 895), 232 (e 7,440), 207 (e 40,994) nm; 1H NMR (DMSO-d6, 300 MHz) 6.34 (s, 2 H), 3.52 (s, 6 H), 3.40 (s, 3 H), 3.11 (t, J = 7.6 Hz, 1 H) 1.76-1.64 (m, 1 H), 1.46-1.36 (m, 1 H), 0.60 (t, J = 7.3 Hz, 3 H); 1H NMR (CD3OD, 300 MHz) 6.78 (s, 2 H), 4.00 (s, 6 H), 3.90 (s, 3 H), 3.55 (t, J = 7.7 Hz, 1 H) 2.24-2.12 (m, 1 H), 1.97-1.83 (m, 1 H), 1.07 (t, J = 7.3 Hz, 3 H); 13C NMR (DMSO-d6, 75 MHz) 175.1, 153.1, 136.9, 135.8, 105.4, 60.3, 56.2, 53.1, 26.7, 12.4; 13C NMR (CD3OD, 75 MHz) 178.1, 154.9, 138.7, 137.4, 106.8, 61.5, 57.0, 55.3, 28.3, 12.9; HRMS (FAB): (M-H)- calcd: 253.1076, meas: 253.1063. Anal. Calcd for C13H18O5: C, 61.41; H, 7.13. Found: C, 61.47; H, 7.20.

[S-(R*,R*)]-1-[1-oxo-2-(3,4,5-trimethoxyphenyl)butyl]-2-piperdinecarboxylic acid (AP17362): A solution of 5 (0.75 g, 2.95 mmol, 91% ee) in CH2Cl2 (15 mL) was treated with methyl-L-pipecolate hydrochloride (0.539 g, 3.00 mmol) followed by 2-chloro-1-methylpyridinium iodide (0.958 g, 3.75 mmol) and triethylamine (1.25 mL, 8.95 mmol). The reaction mixture was allowed to stir for 3.5 h, after which time the solution was diluted with EtOAc (100 mL), washed with water (15 mL), a 5% aqueous citric acid solution (25 mL), a saturated Na2CO3 solution (10 mL), water (15 mL), and finally brine (15 mL). The organic phase was dried over MgSO4 and concentrated to a yellow oil which was then dissolved in MeOH (14 mL). The methanolic solution was treated with water (1 mL) followed by lithium hydroxide monohydrate (0.620 g, 14.78 mmol). After 4 h, the mixture was diluted with EtOAc (100 mL), washed with a saturated NaHCO3 solution (3 x 40 mL) followed by water (20 mL). The aqueous portions were combined and acidified to pH ~3 by careful addition of a 10% aqueous HCl solution. The resulting suspension was extracted with EtOAc (2 x 75 mL) which was then washed with water (2 x 25 mL), brine (20 mL), dried over MgSO4, and concentrated to a solid which was dissolved in a refluxing EtOAc (75 mL) solution and allowed to slowly cool to room temperature. The resulting crystalline material was filtered and air dried to afford product (0.508 g, 42%) as a colorless solid:

(+99% de by Chiralpak AD HPLC with guard column, 0.2:5:95 formic acid/i-PrOH/hexane, retention time 40.0 min for the SR-diastereomer, 43.0 min for the desired SS-diastereomer, 46.5 min for the RR-diastereomer, and 67.5 min for the RS-diastereomer); mp 173.5-174 °C; [a]22D +10.9 (c = 1.01, DMSO, 30 min); UV (MeOH) lmax 270 (e 990), 232 (e 11,161), 207 (e 49,079) nm; 1H NMR (DMSO-

5   d6, 300 MHz) 6.55 (s, 2 H), 5.13 (d, J = 4.4 Hz, 1 H), 3.85-3.64 (m, 11 H), 2.77-2.70 (m, 1 H), 2.12 (d, J = 13.4 Hz, 1 H), 1.99-1.85 (m, 1 H), 1.65-1.55 (m, 4 H), 1.38-1.18 (m, 2 H), 0.84 (t, J = 7.2 Hz, 3 H); 1H NMR (CD3OD, 300 MHz) 6.74 (s, 2 H), 5.43 (d, J = 4.0 Hz, 1 H), 4.13-3.83 (m, 11 H), 3.03 (td, J = 13.5, 3.0 Hz, 1 H), 2.44 (d, J = 13.8 Hz, 1 H), 2.24-2.14 (m, 1 H), 1.90-1.40 (m, 6 H) 1.09 (t, J = 7.3 Hz, 3 H); 13C NMR (DMSO-d6, 75 MHz) 172.9, 172.2, 153.0, 136.2, 105.4, 60.2, 56.2, 56.0, 51.8, 49.4, 43.1,

10   28.5, 26.8, 25.3, 21.0, 12.8; 13C NMR (CD3OD, 75 MHz) 175.4, 174.5, 154.9, 137.5, 106.8, 61.5, 57.1, 53.9, 52.1, 45.2, 29.9, 28.2, 26.8, 22.3, 13.2; HRMS (FAB): (M-H)- calcd: 364.1760, meas: 364.1774. Anal. Calcd for C19H27O6: C, 62.45; H, 7.45; N, 3.83. Found: C, 62.32; H, 7.61; N, 3.88.

15   **Example 4: Regulated activity of a transcription factor**

The chimeric transcription factor ZFHD-p65 consists of the chimeric DNA binding domain, ZFHD1 (Pomerantz et al., Science 267:93-96, 1995)) fused to a transcriptional activation domain from the p65 subunit of NF-kB (Rivera et al., Nat. Med 2:1028-1032, 1996). Transient transfection of the construct into HT1080 cells along with a secreted alkaline phosphatase (SEAP) reporter gene driven

20   by binding sites for ZFHD1 (Rivera et al., Nat. Med 2:1028-1032, 1996) results in the activation of transcription, as measured by the presence of SEAP activity in the culture supernatant (Figure 3). To determine whether the activity of the transcription factor could be made to be dependent on a monomeric ligand, 6 copies of F(36M) were fused to the amino- or carboxy- terminus of the ZFHD-p65 transcription factor. As shown in Figure 3, in the absence of the monomeric ligand, FK506, the activity

25   of the transcription factor is repressed. Treatment of cells with increasing concentration of monomer leads to an increase in the activity of the transcription factor, which peaks at 1 uM. These results suggest that fusion of F(36M) domains to a transcription factor results in its sequestration in an inactive oligomeric complex and that interaction with monomeric ligand results in the release of an active transcription factor.

30

Example 5: Identification of W59V FKBP as a conditional aggregation domain

A second candidate CAD was identified during studies aimed at identifying a second bump-hole system 'orthogonal' to that based around Phe36 mutants. Mutants with truncations of the ligand binding site residue Trp59 (W59) were predicted to be able to bind FKBP

35   ligands that are synthetically modified to bear para-substituents on their pipecolate ring. As described for F36M, fusion proteins were engineered in which three tandem copies of candidate FKBP mutants were fused to the DNA binding domain ZFHD1 and the transcriptional activation domain of the NF-kB p65 subunit (Amara et al., PNAS 94:10618-10623 1997). Mutants were engineered by Kunkel

mutagenesis (Kunkel et al. 1991 Meth. Enzymol. 204: 125-139). The oligonucleotide used to
engineer the mutation W59V is TC-W59V (below). pCGNN expression vectors for the chimeric
proteins were transiently transfected into HT1080L cells that contain a retrovirally integrated SEAP
reporter (Rivera et al., Nature Med 2: 1028-1032 1996).

5

Oligonucleotide TC-W59V: 5' GTGATCCGAGGCgtgGAAGAAGGGGTTGCC

When fusions incorporating the W59V mutant were transfected (Trp mutated to valine),
constitutive transcription SEAP was observed. This transcription could be inhibited by monomeric
10   FKBP ligands such as FK506 or rapamycin, or AP22597, a small synthetic ligand with a para-isopropyl
pipecolate 'bump' designed to recognize the cavity created by the W59V substitution. These data
suggest that the W59V mutant of FKBP self-associates, and that this self-association can be blocked
by suitable FKBP ligands. These data also show that W59V-FKBP fusion proteins could be used as
the basis of a transcriptional "off-switch" to regulate the transcription of target genes.

15   To extend these findings we stably integrated the genes for ZFHD1-(W59V-FKBP)x3 and
(W59V-FKBP)x3-p65 into HT1080L cells, using the procedure described by Amara et al. (PNAS
94:10618-10623 1997). The pooled stably transfected cells constitutively produced SEAP, and this
production could be inhibited in a dose-dependent manner by adding FK506 or AP22597 (see figure
4). Thus protein-protein association between W59V fusions can be detected even when those
20   fusions are expressed at low levels, as in stably transfected cells. These data further demonstrate that
a W59V-based transcriptional off-switch can function in such stably transfected cells.

To further test the aggregation properties of W59V, fusions of more or less than three copies
of the protein can be made to ZFHD1 and p65 and their transcriptional activity assessed by the same
methods described above. Multiple copies of W59V-FKBP can also be fused to EGFP as described
25   for the F36M example. Fluoro-microscopic examination of cells transfected with such constructs will
indicate the extent of aggregation and the number of copies of W59V that provide optimal
aggregation.